

BEST AVAILABLE COPY

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H 15/12, C12N 5/10, 7/02 C12N 7/04, 15/49, C07K 3/12 C07K 13/00, 17/00, C12Q 1/70 A61K 39/10, G01N 33/53		A1	(11) International Publication Number: WO 92/06990 (43) International Publication Date: 30 April 1992 (30.04.92)
(21) International Application Number: PCT/US91/07611 (22) International Filing Date: 17 October 1991 (17.10.91)		(74) Agents: OLIFF, James, A. et al.; Oliff & Berridge, P.O. Box 19928, Alexandria, VA 22320 (US).	
(30) Priority data: 599,491 17 October 1990 (17.10.90) US		(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).	
(71) Applicant: THE UNITED STATES OF AMERICA, represented by THE SECRETARY, UNITED STATES DEPARTMENT OF COMMERCE [US/US]; Washington, DC 20231 (US).		Published <i>With international search report.</i>	
(72) Inventors: REITZ, Marvin, S., Jr. ; 17833 Bowie Mill Road, Derwood, MD 20855 (US). FRANCHINI, Genoveffa ; 4400 17th Street, N.W., Washington, DC 20011 (US). MARKHAM, Phillip, D. ; 17008 Glen Oak Run, Rockville, MD 20855 (US). GALLO, Robert, C. ; 8513 Thornden Terrace, Bethesda, MD 20817 (US). LORI, Franco, C. ; 5517 Southwick Street, Bethesda, MD 20817 (US). POPOVIC, Mikulas ; 9917 Holmhurst Road, Bethesda, MD 20817 (US). GARNTER, Suzanne ; 14512 Cartwright Way, N. Potomac, MD 20878 (US).			

(54) Title: MOLECULAR CLONES OF HIV-1 AND USES THEREOF

(57) Abstract

The present invention relates to the HIV-1 strains MN-ST1 and BA-L which are typical United States HIV-1 isolates. The present invention relates to DNA segments encoding the envelope protein of MN-ST1 or BA-L, to DNA constructs containing such DNA segments and to host cells transformed with such constructs. The viral isolates and envelope proteins of the present invention are of value for use in vaccines and bioassays for the detection of HIV-1 infection in biological samples, such as blood bank samples.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE*	Germany	MC	Monaco	US	United States of America

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

MOLECULAR CLONES OF HIV-1 AND USES THEREOFBACKGROUND OF THE INVENTION

HIV-1 has been identified as the etiologic agent of the acquired immunodeficiency syndrome (AIDS) 5 (Barre-Sinoussi et al., Science 220, 868-871, 1983; Popovic et al., Science 224, 497-500, 1984; Gallo et al., Science 224, 500-503, 1984). Infected individuals generally develop antibodies to the virus within several months of exposure (Sarnagdharan et al., Science 224, 506-508, 10 1984), which has made possible the development of immunologically based tests which can identify most blood samples from infected individuals. This is a great advantage in diagnosis, and is vital to maintaining the maximum possible safety of samples from blood banks.

15 An important aspect of HIV-1 is its genetic variability (Hahn et al., Proc. Natl. Acad. Sci. U.S.A. 82, 4813-4817, 1985). This is particularly evident in the gene for the outer envelope glycoprotein (Starcich et al., Cell 45, 637-648, 1986; Alizon et al., Cell 46, 63-74, 20 1986; Gурго et al., Virology 164, 531-536, 1988). Since the outer envelope glycoprotein is on the surface of the virus particle and the infected cell, it is potentially one of the primary targets of the immune system, including the target of neutralizing antibodies and cytotoxic T 25 cells. This variability may also lead to differences in the ability of antigens from different strains of HIV-1 to be recognized by antibodies from a given individual, as well as to differences in the ability of proteins from different strains of virus to elicit an immune response 30 which would be protective against the mixture of virus strains that exists in the at risk populations.

35 Several biologically active complete molecular clones of various strains of HIV-1 have been obtained and sequenced. These clones, however, seem to represent viral genotypes which are relatively atypical of United States HIV-1 isolates. In addition, several of the translational reading frames for non-structural viral proteins are not complete. Further, viruses derived from these clones do

not grow in macrophages, in contrast to many HIV-1 field isolates and, perhaps, because of this lack of ability to infect macrophages efficiently, these clones do not replicate well in chimpanzees. This latter ability is important for testing candidate vaccines in animal systems. In addition, the ability to infect macrophages is critical in evaluating the possible protective efficacy of elicited immune response since neutralization of infectivity on macrophage may differ from the better studied neutralization on T cells.

Neutralizing antibodies (Robert-Guroff et al., Nature 316, 72-74, 1985; Weiss et al., Nature 316, 69-72, 1985) have been demonstrated in infected individuals, as have cytotoxic T cells responses (Walker et al., Nature 328, 345-348, 1988). Although these do not appear to be protective, it is likely that if they were present prior to infection, they would prevent infection, especially by related strains of virus. This is supported by the finding that macaques can be protected by immunization with inactivated simian immunodeficiency virus (SIV) from infection with the homologous live virus (Murphy-Corb et al., Science 246, 1293-1297, 1989). Chimps also have been passively protected against challenge by live virus by prior administration of neutralizing antibodies to the same virus (Emini et al., J. Virol. 64, 3674-3678, 1989). One problem, however, is that at least some of the neutralizing antibodies studied depend on recognition of a variable region on the envelope (Matsushita et al., J. Virol. 62, 2107-2114, 1988; Rusche et al., Proc. Natl. Acad. Sci. U.S.A. 85, 3198-3202, 1988; Skinner et al., AIDS Res. Hum. Retroviruses 4, 187-197, 1988) called the V3 region (Starcich et al., Cell 45, 637-648, 1986).

An at least partial solution to the problem of viral heterogeneity is to identify prototypical HIV-1 strains, that is, those that are most similar by DNA sequence data or serologic reactivity to strains present in the population at risk. The inclusion of a limited number of such prototype strains in a polyvalent vaccine

cocktail might then result in elicitation of an immune response protective against most naturally occurring viruses within a given population. Such a mixture should also provide the maximum possible sensitivity in diagnostic tests for antibodies in infected individuals.

Components of highly representative isolates of a geographical area provide the maximum possible sensitivity in diagnostic tests and vaccines. Production of viral proteins from molecular clones by recombinant DNA techniques is the preferred and safest means to provide such proteins. Molecular clones of prototype HIV-1 strains can serve as the material from which such recombinant proteins can be made. The use of recombinant DNA avoids any possibility of the presence of live virus and affords the opportunity of genetically modifying viral gene products. The use of biologically active clones ensures that the gene products are functional and hence, maximizes their potential relevance.

Infectious clones, that is, those which after transfection into recipient cells produce complete virus, are desirable for several reasons. One reason is that the gene products are by definition functional; this maximizes their potential relevance to what is occurring in vivo. A second reason is that genetically altered complete virus is easy to obtain. Consequently, the biological consequences of variability can be easily assessed. For example, the effect of changes in the envelope gene on the ability of the virus to be neutralized by antibody can be easily addressed. Using this technique, a single point mutation in the envelope gene has been shown to confer resistance to neutralizing antibody (Reitz et al., Cell 54, 57-63, 1988). A third reason is that a clonal virus population provides the greatest possible definition for challenge virus in animals receiving candidate vaccines, especially those including components of the same molecularly cloned virus.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide vaccine components for an anti HIV-1 vaccine which would represent a typical United States isolate HIV-1.

5 It is another object of the present invention to provide diagnostic tests for the detection of HIV-1.

Various other objects and advantages of the present invention will become apparent from the drawings and the following description of the invention.

10 BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows the structure and restriction map of the lambda MN-PH1 clone.

FIGURE 2 shows the restriction map of the MN-PH1 envelope plasmid clone.

15 FIGURE 3 shows the restriction map and structure of the lambda MN-ST1 clone.

FIGURE 4 shows the structure of the lambda BA-L clone.

20 FIGURE 5 shows the restriction map of the clone BA-L1.

Detailed Disclosure of the Invention

The present invention relates to the HIV-1 virus strains, MN-ST1 and BA-L, which are more typical of the HIV-1 isolates found in the United States than previously 25 known HIV-1 strains. Local isolates provide better material for vaccine and for the detection of the virus in biological samples, such as blood bank samples.

30 The present invention relates to DNA segments encoding the env protein of MN-ST1 or BA-L (the DNA sequence given in Figures 5 and 8 being two such examples) and to nucleotide sequences complementary to the segments referenced above as well as to other genes and nucleotide sequences contained in these clones. The present invention also relates to DNA segments encoding a unique portion of the MN-ST1 env protein or the BA-L env protein. (A "unique portion" consists of at least five (or six) amino acids or corresponding at least 15 (or 18) nucleotides.)

The invention further relates to the HIV-1 virus strains MN-ST1 and BA-L themselves. The HIV-1 virus strains of the present invention are biologically active and can easily be isolated by one skilled in the art using 5 known methodologies.

The above-described DNA segments of the present invention can be placed in DNA constructs which are then used in the transformation of host cells for a generation of recombinantly produced viral proteins. DNA constructs 10 of the present invention comprise a DNA segment encoding the env protein and the flanking region of MN-ST1 (or BA-L) or a portion thereof and a vector. The constructs can further comprise a second DNA segment encoding both a rev protein and a rev-responsive region of the env gene 15 operably linked to the first DNA segment encoding the env protein. The rev protein facilitates efficient expression of the env protein in eucaryotic cells. Suitable vectors for use in the present invention include, but are not limited to, pSP72, lambda EMBL3 and SP65gpt.

Host cells to which the present invention relates 20 are stably transformed with the above-described DNA constructs. The cells are transformed under conditions such that the viral protein encoded in the transforming construct is expressed. The host cell can be procaryotic (such as bacterial), lower eucaryotic (such as fungal, including yeast) or higher eucaryotic (such as mammalian). The host cells can be used to generate recombinantly 25 produced MN-ST1 (or BA-L) env protein by culturing the cells in a manner allowing expression of the viral protein encoded in the construct. The recombinantly produced protein is easily isolated from the host cells using 30 standard protein isolation protocols.

Since HIV-1 strains MN-ST1 and BA-L represent 35 relatively typical United States genotypes, non-infectious MN-ST1 or BA-L proteins (for example, the env protein), peptides or unique portions of MN-ST1 or BA-L proteins (for example, a unique portion of the env protein), and even whole inactivated MN-ST1 or BA-L can be used as an

immunogen in mammals, such as primates, to generate antibodies capable of neutralization and T cells capable of killing infected cells. The protein can be isolated from the virus or made recombinantly from a cloned envelope gene. Accordingly, the virus and viral proteins of the present invention are of value as either a vaccine or a component thereof, or an agent in immunotherapeutic treatment of individuals already infected with HIV-1.

As is customary for vaccines, a non-infectious antigenic portion of MN-ST1 or BA-L, for example, the env protein, can be delivered to a mammal in a pharmacologically acceptable carrier. The present invention relates to vaccines comprising non-infectious antigenic portions of either MN-ST1 or BA-L and vaccines comprising non-infectious antigenic portions of both MN-ST1 and BA-L. Vaccines of the present invention can include effective amounts of immunological adjuvants known to enhance an immune response. The viral protein or polypeptide is present in the vaccine in an amount sufficient to induce an immune response against the antigenic protein and thus to protect against HIV-1 infection. Protective antibodies are usually best elicited by a series of 2-3 doses given about 2 to 3 weeks apart. The series can be repeated when circulating antibody concentration in the patient drops.

Virus derived from the infectious HIV-1(MN) clones, MN-ST1, may also be used for reproducible challenge experiments in chimpanzees treated with candidate HIV-1 vaccines or in vitro with human antiserum from individuals treated with candidate vaccines. A candidate vaccine can be administered to a test mammal, such as a chimpanzee prior to or simultaneously with the infectious MN-ST1 virus of the present invention. Effectiveness of the vaccine can be determined by detecting the presence or absence of HIV-1 infection in the test mammals. Side-by-side comparative tests can be run by further administering to a second set of test mammals the virus alone and comparing the number of infections which develop in the two sets of test mammals. Alternatively, candidate

vaccines can be evaluated in humans by administering the vaccine to a patient and then testing the ability of the MN-ST1 virus to infect blood cells from the patient.

The present invention also relates to the detection of HIV-1 virus in a biological sample. For detection of an HIV-1 infection, the presence of the virus, proteins encoded in the viral genome, or antibodies to HIV-1 is determined. Many types of tests, as one skilled in the art will recognize, can be used for detection. Such tests include, but are not limited to, ELISA and RIA.

In one bioassay of the present invention all, or a unique portion, of the env protein is coated on a surface and contacted with the biological sample. The presence of a resulting complex formed between the protein and antibodies specific therefor in the serum can be detected by any of the known methods commonly used in the art, such as, for example, fluorescent antibody spectroscopy or colorimetry.

The following non-limiting examples are given to further demonstrate the present invention without being deemed limitative thereof.

EXAMPLES

MN-PH1 Clone

The permuted circular unintegrated viral DNA representing the complete HIV-1(MN) genome was cloned by standard techniques (Sambrook et al., 1989, Molecular Cloning. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press) into the Eco RI site of lambda gtWES.lambdab DNA from total DNA of H9 cells producing HIV-1(MN). This clone is designated lambda MN-PH1, and its structure and restriction map are shown in Figure 1. The clone was subcloned into M13mp18 and M13mp19, and the DNA sequence of the entire clone, given in Figure 2, was obtained by the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467, 1977). The amino acid sequence of the envelope protein (see Table I) was inferred from the DNA sequence. A restriction map of the cloned unintegrated viral DNA (see

Figure 1) was also obtained from the DNA sequence of lambda PH1 and used in conjunction with the inferred amino acid sequence of the viral proteins to subclone the envelope (env) gene into the commercially available 5 plasmid pSP72 (Promega Biological Research Products, Madison, WI), as shown in Figure 2. This plasmid (pMN- PH1env) contains, in addition to the coding regions for the envelope proteins, the coding region for the rev protein (Feinberg et al., Cell 46, 807-817, 1986) and the 10 portion of the env gene which contains the rev-responsive region (Dayton et al., J. Acquir. Immune. Defic. Syndr. 1, 441-452, 1988), since both are necessary for efficient expression of the envelope protein in eucaryotic cells. This plasmid thus contains all the elements required for 15 production of envelope protein following placement into appropriate expression vectors and introduction into recipient cells, all by standard techniques known to molecular biologists.

MN-ST1 Clone

20 The infectious molecular clone, lambda MN-ST1, was obtained by cloning integrated provirus from DNA purified from peripheral blood lymphocytes infected with HIV-1(MN) and maintained in culture for a short time (one month). The integrated proviral DNA was partially digested with 25 the restriction enzyme Sau3A under conditions which gave a maximum yield of DNA fragments of from 15-20 kilobases (kb). This was cloned into the compatible BamHI site of lambda EMBL3, as shown in Figure 3. Figure 3 also shows the restriction map of clone lambda MN-ST1. The DNA 30 sequence of the entire clone, given in Table II, was obtained by the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467, 1977). The amino acid sequence was predicted from the DNA sequence (see Table II). This clone can be transfected 35 into recipient cells by standard techniques. After transfection, the cloned proviral DNA is expressed into biologically active virus particles, which can be used as a source for virus stocks. The proviral DNA whose

restriction map is shown in Figure 2, was removed from the lambda phage vector by digestion with BamHI and inserted into a plasmid, SP65gpt (Feinberg et al., Cell 46, 807-817, 1986). This plasmid, pMN-ST1, contains an SV40 5 origin of replication. Consequently, transfection into COS-1 cells (Gluzman, Y. Cell 23, 175-182, 1981), which produce a SV40 gene product which interacts with the cognate origin of replication, results in a transient high 10 plasmid copy number with a concomitant production of large amount of replication competent, infectious virus (Feinberg et al., Cell 46, 807-817, 1986). This provides a convenient source of genetically homogeneous virus, as well as a way to introduce desired mutations using standard methods.

15 The envelope gene was excised from the lambda phage clone and cloned into a plasmid as described above for lambda MN-PH1. This clone (pMN-ST1env), is similar to pMN-PH1env, described above, except that it derives from a biologically active cloned provirus. Like pMN-PH1env, it 20 can be placed in a suitable vector and host to produce the envelope protein of HIV-1(MN) by well known techniques.

BA-L Clone

25 A Hind III fragment of unintegrated viral DNA representing the HIV-1(BA-L) genome was cloned by standard techniques into lambda phage Charon 28 DNA from total DNA of peripheral blood macrophages infected with and producing HIV-1(BA-L). A positive clone was selected by hybridization using a radiolabelled probe for the HIV-1 envelope. This clone, designated lambda BA-L1, was found 30 to contain the entire gene for the envelope protein. Its structure is given in Figure 4. The insert was transferred into a plasmid (pBluescript, Stratagene, LaJolla, CA) and the DNA sequence of the env gene was determined (see Table III). This clone is designated pBA-L1.

35 The amino acid sequence of the envelope protein, shown in Table III, was inferred from the DNA sequence. A restriction map was also obtained from the DNA sequence of BA-L1 (shown in Figure 5) in order to determine the

appropriate restriction enzyme sites for cloning the env gene into suitable expression vectors. An Eco RI-HindIII fragment of 0.4 Kb and a 2.8 Kb HindIII-XbaI fragment when cloned together constitute the entire env gene. This 5 plasmid contains, in addition to the coding regions for the envelope proteins, the coding region for the rev protein and the portion of the env protein which contains the rev-responsive region. Both are necessary for efficient expression of the envelope protein in eucaryotic 10 cells (Feinberg et al., Cell 46, 807-817, 1986; Dayton et al., J. Acquir. Immune. Defic. Syndr. 1, 441-452). This plasmid thus contains all the HIV-1 genetic elements required for production of envelope protein following placement into appropriate expression vectors and introduction into recipient cells, all by standard techniques 15 well known in the art.

Statement of Deposit

The lambda MN-ST1 clone and the BA-L plasmid clone were deposited at the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., on September 13, 1990, under the terms of the Budapest Treaty. The lambda MN-ST1 clone has been assigned the ATCC accession number ATCC 40889 and the BA-L plasmid clone has been assigned the ATCC accession number 25 ATCC 40890.

* * * * *

All publications mentioned hereinabove are hereby incorporated by reference.

While the foregoing invention has been described 30 in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention.

TABLE I

TGGAAGGGCT AATTCACTCC CAACGAAGAC AAGATATCCT	TGATCTGTGG ATCTACCACA	60
CACAAGGCTA CTTCCCTGAT TAGCAGAACT ACACACCAGG	GCCAGGGATC AGATATCCAC	120
TGACCTTTGG ATGGTGCTAC AAGCTAGTAC CAGTTGAGCC	AGAGAAGTTA GAAGAAGCCA	180
ACAAAGGAGA GAACACCAGC TTGTTACACC CTGTGAGCCT	GCATGGAATG GATGACCCGG	240
AGAGAGAAGT GTTAGAGTGG AGGTTGACA GCCGCCTAGC	ATTTCATCAC ATGGCCCGAG	300
AGCTGCATCC GGAGTACTTC AAGAACTGCT GACATCGAGC	TTGCTACAAG GGACTTTCCG	360
CTGGGGACTT TCCAGGGAGG CGTGGCCTGG GCGGGACTGG	GGAGTGGCGA GCCCTCAGAT	420
CCTGCATATA AGCAGCTGCT TTTGCTGT ACTGGGTCTC	TCTGGTTAGA CCAGATCTGA	480
GCCTGGGAGC TCTCTGGCTA ACTAGGGAAC CCACTGCTTA	AGCCTCAATA AAGCTTGCCT	540
TGAGTGCTTC AAGTAGTGTG TGCCCGTCTG TTATGTGACT	CTGGTAGCTA GAGATCCCTC	600
AGATCCCTTT AGGCAGTGTG GAAAATCTCT AGCAGTGGCG	CCCCAACAGG GACTTGAAAG	660
CGAAAGAAAA ACCAGAGCTC TCTCGACGCA GGACTCGGCT	TGCTGAAGCG CGCACGGCAA	720
GAGGCGAGGG GCGGCGACTG GTGAGTACGC CAAAAATTCT	TGACTAGCGG AGCCTAGAAG	780
GAGAGAGATG GGTGCGAGAG CGTCGGTATT AAGCGGGGGA	GAATTAGATC GATGGGAAAA	840
CATTGGTTA AGGCCAGGGG GAAAGAAAAA ATATAAATTAA	AAACATGTAG TATGGGCAAG	900
CAGGGAGCTA GAACGATTG CAGTCAATCC TGGCCTGTTA	GAACATCAG AAGGCTGTAG	960
ACAAATACTG GGACAGCTAC AACCATCCCT TCAGACAGGA	TCAGAAGAAC TAAATCATT	1020
ATATAATACA GTAGCAACCC TCTATTGTGT GCATCAAAG	ATAGAGATAA AAGACACCAA	1080
GGAAAGCTTTA GAGAAAATAG AGGAAGAGCA AAACAAAAGT	AAGAAAAAAG CACAGCAAGC	1140
AGCAGCTGAC ACAGGAAACA GAGGAAACAG CAGCCAAGTC	AGCCAAAATT ACCCCATAGT	1200
GCAGAACATC GAGGGGCAA TGTCATCA GGCCATATCA	CCTAGAACTT TAAATGCATG	1260
GGTAAAAGTA GTAGAAGAGA AGGCTTCAG CCCAGAAGTA	ATACCCATGT TTTCAGCATT	1320
ATCAGAAGGA GCCACCCAC AAGATTAAA CACCATGCTA	AAACACAGTGG GGGGACATCA	1380
AGCAGCCATG CAAATGTTAA AAGAGACCAT CAATGAGGAA	GCTGCAGAAT GGGATAGATT	1440
GCATCCAGTG CATGCAGGGC CTATTACACC AGGCCAGATG	AGAGAACCAA GGGGAAGTGA	1500
CATAGCAGGA ACTACTAGTA CCCTTCAGGA ACAAAATAGGA	TGGATGACAA ATAATCCACC	1560
TATCCCAGTA GGAGAAAATCT ATAAAAGATG GATAATCCTG	GGATTTAAATA AAATAGTAAG	1620
GATGTATAGC CCTTCCAGCA TTCTGGACAT AAGACAAGGA	CCAAAGGAAC CCTTTAGAGA	1680
CTATGTAGAC CGGTTCTATA AAACTCTAAG AGCCGAGCAA	GCTTCACAGG AGGTAAAAAA	1740
CCGGACGACA GAAACCTTGT TGGTCCAAA TGCGAACCCA	GATTGTAAGA CTATTTAAA	1800
AGCATTGGGA CCAGCAGCTA CACTAGAAGA AATGATGACA	GCATGTCAGG GAGTGGGAGG	1860
ACCTGGTCAT AAAGCAAGAG TTTTGGCGGA AGCGATGAGC	CAAGTAACAA ATTCACTAC	1920

CATAATGATG CAGAGAGGCA ATTTAGGAA TCAAAGAAG ATTATCAAGT GCCTCAATTG 1980
TGGCAAAGAA GGGCACATAG CCAAAAATTG CAGGGCCCCCT AGGAAAAGGG GCTGTTGGAA 2040
ATGTGAAAG GAAGGACACC AAATGAAAGA TTGTACTGAG AGACAGGCTA ATTTTTTAGG 2100
GAAGATCTGG CCTTCCTGCA AGGGAAAGCG GAATTTCCCT CAGAGCAGAA CAGAGCCAAC 2160
AGCCCCACCA GAAGAGAGCT TCAGGTTGG GGAAGAGACA ACAACTCCCT ATCAGAAGCA 2220
GGAGAAGAAG CAGGAGACGA TAGACAAGGA CCTGTATCCT TTAGCTTCCC TCAAATCACT 2280
CTTTGGCAAC GACCCATTGT CACAATAAG ATAGGGGGGC AACTAAAGGA AGCTCTATTA 2340
GATACAGGAG CAGATGATAC AGTATTAGGA GAAATGAATT TGCCAAGAAG ATGAAACCCA 2400
AAAATGATAG GGGGAATTGG AGGTTTTATC AAAGTAAGAC AGTATGATCA GATAACCATA 2460
GGAATCTGTG GACATAAAAGC TATAGGTACA GTATTAGTAG GACCTACACC TGTCAACATA 2520
ATTGGAAGAA ATCTGTTGAC TCAGCTTGGG TGCACTTAA ATTTCCCAT TAGTCCTATT 2580
GAAACTGTAC CAGTAAAATT AAAGCCAGGA ATGGATGGCC CAAAAGTTAA ACAATGGCCA 2640
TTGACAGAAG AAAAAATAAA AGCATTAATA GAAATTGTA CAGAAATGGA AAAGGAAGGG 2700
AAAATTCAA AAATTGGGCC TGAAAATCCA TACAATACTC CAGTATTGTC CATAAGAAA 2760
AAAGACAGTA CTAAATGGAG AAAATTAGTA GATTTCAGAG AACTTAATAA GAAAACCTCAA 2820
GACTTCTGGG AAGTTCAATT AGGAATACCA CATCCTGCAG GGTTAAAAAA GAAAAAAATCA 2880
GTAACAGTAC TGGATGTTGGG TGATGCATAT TTTTCAGTTC CCTTAGATAA AGACTTCAGG 2940
AAGTATACTG CATTACCAT ACCTAGTATA ACAATGAAA CACCAGGGAT TAGATATCAG 3000
TACAATGTGC TTCCACAGGG ATGGAAAGGA TCACCAGCAA TATTCCAAAG TAGCATGACA 3060
AAAATCTTAG AGCCTTTAG AAAACAAAAT CCAGACATAG TTATCTATCA ATACATGGAT 3120
GATTTGTATG TAGGATCTGA CTTAGAAATA GGGCAGCATA GAGCAAAAT AGAGGAACGT 3180
AGACGACATC TGTTGAGGTG GGGATTTACC ACACCAAGACA AAAACATCA GAAAGAACCT 3240
CCATTCTTT GGATGGGTTA TGAACCTCAT CCTGATAAAAT GGACAGTACA GCCTATAGTG 3300
CTACCAAGAAA AAGACAGCTG GACTGTCAAT GACATACAGA AGTTAGTGGG AAAATTGAAT 3360
TGGGCAAGTC AGATTTACGC AGGGATTAAA GTAAAGCAAT TATGTAACCT CCTTAGAGGA 3420
ACCAAAGCAC TAACAGAACT AATACCACTA ACAGAAGAAG CAGAGCTAGA ACTGGCAGAA 3480
AACAGGGAAA TTCTAAAAGA ACCAGTACAT GGAGTGTATT ATGACCCATC AAAAGACTTA 3540
ATAGCAGAAG TACAGAAGCA GGGGCAAGGC CAATGGACAT ATCAAATTAA TCAAGAGCCA 3600
TTTAAAAATC TGAAAACAGG CAAATATGCA AGAATGAGGG GTGCCACAC TAATGATGTA 3660
AAACAATTAA CAGAGGCAGT GCAAAAAATA GCCACAGAAA GCATAGTAAT ATGGGGAAAG 3720
ACTCCTAAAT TTAGACTACC CATAACAAAAA GAAACATGGG AAACATGGTG GACAGAGTAT 3780
ACGTAAGCCA CCTGGATTCC TGAGTGGGAG GTTGTCAATA CCCCTCCCTT AGTGAATTA 3840
TGGTACCAAGT TAGAGAAAGA ACCCATAGTA GGTGCAGAAA CTTTCTATGT AGATGGGCA 3900
GCTAACAGGG AGACTAAAAA AGGAAAAGCA GGATATGTTA CTAACACAGG AAGACAAAAG 3960

GTTGTCTCCC TAACTGACAC AACAAATCAG AAGACTGAGT TACAAGCAAT TCATCTAGCT 4020
TTGCAAGATT CAGGGTTAGA AGTAAACATA GTAACAGACT CACAATATGC ATTAGGAATC 4080
ATTCAAGCAC AACCAAGATAA AAGTGAATCA GAGTTAGTCA GTCAAATAAT AGAGCAGTTA 4140
ATAAAAAAAGG AAAAGGTCTA TCTGGCATGG GTACCAGCAC ACAAAAGGAAT TGGAGGAAAT 4200
GAACAAGTAG ATAAATTAGT CAGTGCTGGA ATCAGGAAAG TACTATTTT AGATGGAATA 4260
GATAAGGCCA AAGAAGACCA TGAGAAATAT CACAGTAATT GGAGAGCAAT GGCTAGTGAC 4320
TTAACCTAC CACCTATACT AGCAAAAGAA ATAGTAGCCA GCTGTGATAA ATGTCAGCTA 4380
AAAGGAGAAG CCATGCATGG ACAACTAGAC TGAGTCCAG GAATATGGCA ACTAGATTGT 4440
ACACATTTAG AAGGAAAAGT TATCCTGGTA GCAGTTCATG TAGCCAGTGG ATACATAGAA 4500
GCAGAAGTTA TTCCAGCAGA GACAGGGCAG GAGACAGCAT ACTTTCTCTT AAAATTAGCA 4560
GGAAGATGGC CAGTAAAAC AATACATACA GACAATGGCC CCAATTTCAC CAGTACTACG 4620
GTTAAGGCCG CCTGTTGGTG GACGGGAATC AAGCAGGAAT TTGGCATTCC CTACAATCCC 4680
CAAAGTCAAG GAGTAATAGA ATCTATGAAT AAAGAATTAA AGAAAATTAT AGGACAGGTA 4740
AGAGATCAGG CTGAACATCT TAAGAGAGCA GTACAAATGG CAGTATTCTAT CCACAATTTT 4800
AAAAGAAAAG GGGGGATTGG GGGGTACAGT GCAGGGGAAA GAATAGTAGG CATAATAGCA 4860
ACAGACATAC AAACTAAAGA ACTACAAAAA CAAATTACAA AAATTCAAAA TTTTCGGGTT 4920
TATTACAGGG ACAGCAGAGA TCCACTTTGG AAAGGACCAAG CAAAGCTTCT CTGGAAAGGT 4980
GAAGGGGCAG TAGTAATACA AGATAATAAT GACATAAAAG TAGTGCCAAG AAGAAAAGCA 5040
AAGGTCTTAA GGGATTATGG AAAACAGACG GCAGGTGATG ATTGTGTGGC AAGCAGACAG 5100
GATGAGGATT AGAACATGGA AAAGTTAGT AAAACACCCT ATGTATATTT CAAAGAAAAGC 5160
TAAAGGACGG TTTTATAGAC ATCACTATGA AAGCACTCAT CCAAGAATAA GTTCAGAAGT 5220
ACACATCCCA CTAGGGGATG CTAGATTGGT AATAACAACA TATTGGGTC TGCATACAGG 5280
AGAAAGAGAC TGGCATTAG GTCAGGGAGT CTCCATAGAA TGGAGGAAAA AGAGATATAG 5340
CACACAAGTA GACCCTGACC TAGCAGACCA CCTAATTCTAT CTGCATTACT TTGATTGTTT 5400
TTCAGACTCT GCCATAAGAA AGGCCATATT AGGACATAGA GTTAGTCCTA TTTGTGAATT 5460
TCAAGCAGGA CATAACAAGG TAGGACCTCT ACAGTACTTG GCACAAACAG CATTAAATAAC 5520
ACCAAAAAAG ATAAAGCCAC CTTTGCCTAG TGTTAAGAAA CTGACAGAGG ATAGATGGAA 5580
CAAGCCCCAG AAGACCAAGG GCCACAGAGG GAGCCATACA ATCAATGGC ACTAGAGCTT 5640
TTAGAGGAGC TTAAGAATGA AGCTGTTAGA CATTTCCTA GGATATGGCT CCATGGCTTA 5700
GGGCAACATA TCTATGAAAC TTATGGGAT ACTTGGGCAG GAGTGGAAAGC CATAATAAGA 5760
ATTCTACAAAC AACTGCTGTT TATTCAATTG AGAATTGGGT GTCGACATAG CAGAATAGGC 5820
ATTATTCGAC AGAGGAGAGC AAGAAATGGA GCCAGTAGAT CCTAGACTAG AGCCCTGGAA 5880
GCATCCAGGA AGTCAGCCTA AGACTGCTTG TACCACTTGC TATTGTAAAA AGTGTGCTT 5940

TCATTGCCAA GTTTGTTCAGAAAAAAAGC CTTAGGCATC TCCTATGGCA GGAAGAAGCG 6000
 GAGACAGCGA CGAAGAGCTC CTGAAGACAG TCAGACTCAT CAAGTTCTC TACCAAAGCA 6060
 GTAAAGTAGTA CATGTAATGC AACCTTAGT AATAGCAGCA ATAGTAGCAT TAGTAGTAGC 6120
 AGGAATAATA GCAATAGTTG TGTGATCCAT AGTATTCCATA GAATATAGGA AAATAAGAAG 6180
 ACAAAAGAAAA ATAGACAGGT TAATTGATAG AATAAGCGAA AGAGCAGAAG ACAGTGGCA 6239
 ATG AGA GTG AAG GGG ATC AGG AGG AAT TAT CAG CAC TGG TGG GGA TGG 6287
 Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp
 1 5 10 15
 GGC ACG ATG CTC CTT GGG TTA TTA ATG ATC TGT AGT GCT ACA GAA AAA 6335
 Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu Lys
 20 25 30
 TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT GTG TGG AAA GAA GCA ACC 6383
 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
 35 40 45
 ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA GCA TAT GAT ACA GAG GTA 6431
 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
 50 55 60
 CAT AAT GTT TGG GCC ACA CAA GCC TGT GTA CCC ACA GAC CCC AAC CCA 6479
 His Asn Val Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80
 CAA GAA GTA GAA TTG GTA AAT GTG ACA GAA AAT TTT AAC ATG TGG AAA 6527
 Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp Lys
 85 90 95
 AAT AAC ATG GTA GAA CAG ATG CAT GAG GAT ATA ATC AGT TTA TGG GAT 6575
 Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110
 CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC CCA CTC TGT GTT ACT TTA 6623
 Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125
 AAT TGC ACT GAT TTG AGG AAT ACT ACT AAT ACC AAT AAT AGT ACT GCT 6671
 Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Asn Ser Thr Ala
 130 135 140
 AAT AAC AAT AGT AAT AGC GAG GGA ACA ATA AAG GGA GGA GAA ATG AAA 6719
 Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met Lys
 145 150 155 160
 AAC TGC TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT AAG ATG CAG AAA 6767
 Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Lys Met Gln Lys
 165 170 175
 GAA TAT GCA CTT CTT TAT AAA CTT GAT ATA GTA TCA ATA GAT AAT GAT 6815
 Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asp Asn Asp
 180 185 190
 AGT ACC AGC TAT AGG TTG ATA AGT TGT AAT ACC TCA GTC ATT ACA CAA 6863
 Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln
 195 200 205
 GCT TGT CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAC TAT TGT GCC 6911
 Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
 210 215 220

CCG GCT GGT TTT GCG ATT CTA AAA TGT AAC GAT AAA ARG TTC AGT GGA 6959
 Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser Gly
 225 230 235 240

AAA GGA TCA TGT AAA AAT GTC AGC ACA GTC CAA TGT ACA CAT GGA ATT 7007
 Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
 245 250 255

AGG CCA GTA GTA TCA ACT CAA CTG CTG TTA AAT GGC AGT CTA GCA GAA 7055
 Arg Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala Glu
 260 265 270

GAA GAG GTA GTA ATT AGA TCT GAG AAT TTC ACT GAT AAT GCT AAA ACC 7103
 Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr
 275 280 285

ATC ATA GTA CAT CTG AAT GAA TCT GTA CAA ATT AAT TGT ACA AGA CCC 7151
 Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro
 290 295 300

AAC TAC AAT AAA AGA AAA AGG ATA CAT ATA GGA CCA GGG AGA GCA TTT 7199
 Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala Phe
 305 310 315 320

TAT ACA ACA AAA AAT ATA ATA GGA ACT ATA AGA CAA GCA CAT TGT AAC 7247
 Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys Asn
 325 330 335

ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AGA CAG ATA GTT ACC AAA 7295
 Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser Lys
 340 345 350

TTA AAA GAA CAA TTT AAG AAT AAA ACA ATA GTC TTT AAT CAA TCC TCA 7343
 Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser Ser
 355 360 365

GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA GGG GAA 7391
 Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
 370 375 380

TTT TTC TAC TGT AAT ACA TCA CCA CTG TTT AAT AGT ACT TGG AAT GGT 7439
 Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly
 385 390 395 400

AAT AAT ACT TGG AAT AAT ACT ACA GGG TCA AAT AAC AAT ATC ACA CTT 7487
 Asn Asn Thr Trp Asn Asn Thr Gly Ser Asn Asn Asn Ile Thr Leu
 405 410 415

CAA TGC AAA ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA AAA 7535
 Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys
 420 425 430

GCA ATG TAT GCC CCT CCC ATT GAA GGA CAA ATT AGA TGT TCA TCA AAT 7583
 Ala Met Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser Asn
 435 440 445

ATT ACA GGG CTA CTA TTA ACA AGA GAT GGT AAG GAC ACG GAC ACG 7631
 Ile Thr Gly Leu Leu Leu Arg Asp Gly Gly Lys Asp Thr Asp Thr
 450 455 460

AAC GAC ACC GAG ATC TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT 7679
 Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp Asn
 465 470 475 480

TGG AGA AGT GAA TTA TAT AAA TAT AAA GCA GCA ACA ATT GAA CCA TTA Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu 485 490 495	7727
GGA GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAG AGA GAA AAA Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys 500 505 510	7775
AGA GCA GCG ATA GGA GCT CTG TTC CTT GGG TTC TTA GGA GCA GCA GGA Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly 515 520 525	7823
AGC ACT ATG GGC GCA GCG TCA GTG ACG CTG ACG GCA CAG GCC AGA CTA Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu 530 535 540	7871
TTA TTG TCT GGT ATA GTG CAA CAG CAG AAC AAT TTG CTG AGG GCC ATT Leu Leu Ser Gly Ile Val Gln Gln Asn Asn Leu Leu Arg Ala Ile 545 550 555 560	7919
GAG GCG CAA CAG CAT ATG TTG CAA CTC ACA GTC TGG GGC ATC AAG CAG Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln 565 570 575	7967
CTC CAG GCA AGA GTC CTG GCT GTG GAA AGA TAC CTA AAG GAT CAA CAG Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln 580 585 590	8015
CTC CTG GGG TTT TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC ACT ACT Leu Leu Gly Phe Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Thr 595 600 605	8063
GTG CCT TGG AAT GCT AGT TGG AGT AAT AAA TCT CTG GAT GAT ATT TGG Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile Trp 610 615 620	8111
AAT AAC ATG ACC TGG ATG CAG TGG GAA AGA GAA ATT GAC AAT TAC ACA Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr 625 630 635 640	8159
AGC TTA ATA TAC TCA TTA CTA GAA AAA TCG CAA ACC CAA CAA GAA AAG Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu Lys 645 650 655	8207
AAT GAA CAA GAA TTA TTG GAA TTG GAT AAA TGG GCA AGT TTG TGG AAT Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn 660 665 670	8255
TGG TTT GAC ATA ACA AAT TGG CTG TGG TAT ATA AAA ATA TTC ATA ATG Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met 675 680 685	8303
ATA GTA GGA GGC TTG GTA GGT TTA AGA ATA GTT TTT GCT GTA CTT TCT Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser 690 695 700	8351
ATA GTG AAT AGA GTT AGG CAG GGA TAC TCA CCA TTG TCG TTG CAG ACC Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr 705 710 715 720	8399
CGC CCC CCA GTT CCG AGG GGA CCC GAC AGG CCC GAA GGA ATC GAA GAA Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu 725 730 735	8447

GAA GGT GGA GAG AGA GAC AGA GAC ACA TCC GGT CGA TTA GTG CAT GGA	8495
Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His Gly	
740	745
750	
TTC TTA GCA ATT ATC TGG GTC GAC CTG CGG AGC CTG TTC CTC TTC AGC	8543
Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser	
755	760
765	
TAC CAC CAC AGA GAC TTA CTC TTG ATT GCA GCG AGG ATT GTG GAA CTT	8591
Tyr His His Arg Asp Leu Leu Ile Ala Ala Arg Ile Val Glu Leu	
770	775
780	
CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT CTC CTA	8639
Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu	
785	790
795	800
CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG CTT AAT	8687
Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu Asn	
805	810
815	
GCC ACA GCT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA GAA GTA	8735
Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val	
820	825
830	
CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA ATA AGA	8783
Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg Ile Arg	
835	840
845	
CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT CAAAACGTGT	8837
Gln Gly Leu Glu Arg Ala Leu Leu	
850	855
GAATGGATGG CCTACTGTAA GGGAAAGAAT GAGACGAGCT GAACCAGCTG AGCTAGCAGC	8897
AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA CAAGTAGCAA	8957
TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAGAGGAGG AGGAAGTGGG	9017
TTTCCAGTC AACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG CTTAGATCT	9077
TAGCCACTTT TAAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC AAAAGAGACA	9137
AGACATCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT GGCAGAACTA	9197
CACACCAGGG CCAGGGATCA GATATCCACT GACCTTTGGA TGGTGCTTCA AGCTAGTACC	9257
AGTTGAGCCA GAGAACATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT TGTACACCCC	9317
TATGAGCCAG CATGGATGGA TGACCCGGAG AGAGAAAGTGT TAGTGTGGAA GTCTGACAGC	9377
CACCTAGCAT TTCAGCATTG TGCCCGAGAG CTGCATCCGG AGTACTACAA GAACTGCTGA	9437
CATCGAGCTA TCTACAAGGG ACTTCCGCT GGGGACTTTC CAGGGAGGTG TGGCTGGC	9497
GGGACCGGGG AGTGGCGAGC CCTCAGATCG TGCATATAAG CAGCTGCTTT CTGCCTGTAC	9557
TGGGTCTCTC TGGTTAGACC AGATCTGAGC CTGGGAGCTC TCTGGCTAAC TAGGAAACCC	9617
ACTGCTTAAG CCTCAATAAA GCTTGCCCTTG AGTGCTCAA GTAGTGTGTG CCCGTCTGTT	9677
ATGTGACTCT GGTAGCTAGA GATCCCTCAG ATCCTTTAG GCAGTGTGGA AAATCTCTAG	9737
CA	9739
Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp	
1	5
10	15

Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu Lys
 20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
 35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
 50 55 60

His Asn Val Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80

Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp Lys
 85 90 95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125

Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Ser Thr Ala
 130 135 140

Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met Lys
 145 150 155 160

Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Lys Met Gln Lys
 165 170 175

Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asp Asn Asp
 180 185 190

Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln
 195 200 205

Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
 210 215 220

Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser Gly
 225 230 235 240

Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
 245 250 255

Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu
 260 265 270

Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr
 275 280 285

Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro
 290 295 300

Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala Phe
 305 310 315 320

Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys Asn
 325 330 335

Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser Lys
 340 345 350

Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser Ser
 355 360 365

Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
 370 375 380
 Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly
 385 390 395 400
 Asn Asn Thr Trp Asn Asn Thr Thr Gly Ser Asn Asn Asn Ile Thr Leu
 405 410 415
 Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys
 420 425 430
 Ala Met Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser Asn
 435 440 445
 Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp Thr
 450 455 460
 Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp Asn
 465 470 475 480

 Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu
 485 490 495
 Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
 500 505 510
 Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly
 515 520 525
 Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu
 530 535 540
 Leu Leu Ser Gly Ile Val Gln Gln Asn Asn Leu Leu Arg Ala Ile
 545 550 555 560
 Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln
 565 570 575
 Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln
 580 585 590
 Leu Leu Gly Phe Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Thr
 595 600 605
 Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile Trp
 610 615 620
 Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr
 625 630 635 640
 Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu Lys
 645 650 655
 Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn
 660 665 670

 Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met
 675 680 685
 Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser
 690 695 700
 Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr
 705 710 715 720

Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu
725 730 735

Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His Gly
740 745 750

Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser
755 760 765

Tyr His His Arg Asp Leu Leu Ile Ala Ala Arg Ile Val Glu Leu
770 775 780

Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu
785 790 795 800

Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu Asn
805 810 815

Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val
820 825 830

Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg Ile Arg
835 840 845

Gln Gly Leu Glu Arg Ala Leu Leu
850 855

TABLE II

TGGATGGTT AATTTACTCC CAAAGAGACA AGACATCCTT GATCTGTGGG TCTACCACAC	60
ACAAGGCTAC TTCCCTGATT GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT	120
GACCTTTGGA TGGTGCTTCA AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA	180
TAAAGGAGAG AAACAACGTCT TGTACACCC TATGAGCCAG CATGGATGG ATGACCCGGA	240
GAGAGAAGTG TTAGTGTGGA AGTCTGACAG CCACCTAGCA TTCAGCATT ATGCCCGAGA	300
GCTGCATCCG GAGTACTACA AGAACTGCTG ACATCGAGCT ATCTACAGG GACTTTCCGC	360
TGGGGACTTT CCAGGGAGGT GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG	420
CTGCATATAA GCAGCTGCTT TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG	480
CCTGGGAGCT CTCTGGCTAA CTAGGAAACC CACTGCTAA GCCTCAATAA AGCTTGCCCTT	540
GAGTGCTTCA AGTAGTGTGT GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA	600
GATCCTTTA GGCAGTGTGG AAAATCTCTA GCAGTGGCGC CCGAACAGGG ACTTGAAAGC	660
GAAAGAGAAA CCAGAGGAGC TCTCTCGACG CAGGACTCGG CTTGCTGAAG CGCCACCGC	720
AAGAGGCGAG GGGCGGGGAC TGGTAGTAC GCCAAATTG TTGACTAGCG GAGGCTAGAA	780
GGAGAGAGAT GGGTGCAGAGA GCGTCGGTAT TAAGCGGGGG AGAATTAGAT CGATGGGAAA	840
AAATTGGTT AAGGCCAGGG GGAAAGAAAA AATATAAATT AAAACATGTA GTATGGCAA	900
GCAGGGAGCT AGAACGATTC GCAGTCAATC CTGGCCTGTT AGAAACATCA GAAGGCTGTA	960
GACAAATACT GGGACAGCTA CAACCATCCC TTCAGACAGG ATCAGAAGAA CTTAAATCAT	1020
TATATAATAC AGTAGCAACC CTCTATTGTG TGCATAAAA GATAGAGATA AAAGACACCA	1080
AGGAAGCTTT AGAGAAAATA GAGGAAGAGC AAAACAAAAG TAAGAAAAAA GCACAGCAAG	1140
CAGTAGCTGA CACAGGAAAC AGAGGAAACA GCAGCCAAGT CAGCCAAAT TACCCCATAG	1200
TGCAGAACAT CCAGGGCAA ATGGTACATC AGGCCATATC ACCTAGAACT TTAAATGCAT	1260
GGGTAAAAGT AGTAGAAGAG AAGGCTTTCA GCCCAGAAGT AATACCCATG TTTTCAGCAT	1320
TATCAGAAGG AGCCACCCCA CAAGATTAA ACACCATGCT AAACACAGTG GGGGGACATC	1380
AAGCAGCCAT GCAAATGTTA AAAGAGACCA TCAATGAGGA AGCTGCAGAA TGGGATAGAT	1440
TGCATCCAGT GCATGCAGGG CCTATTGCAC CAGGCCAGAT GAGAGAACCA AGGGGAAGTG	1500
ACATAGCAGG AACTACTAGT ACCCTTCAGG AACAAATAGG ATGGATGACA AATAATCCAC	1560
CTATCCCAGT AGGAGAAAATC TATAAAAGAT GGATAATCCT GGGATTTAAAT AAAATAGTAA	1620
GGATGTATAG CCCTTCCAGC ATTCTGGACA TAAGACAAGG ACCAAAGGAA CCCTTTAGAG	1680
ACTATGTAGA CCGGTTCTAT AAAACTCTAA GAGCCGAGCA AGCTTCACAG GAGGTAAAAA	1740
ATTGGATGAC AGAAACCTTG TTGGTCCAAA ATGCGAACCC AGATTGTAAG ACTATTTAA	1800
AAGCATTGGG ACCAGCAGCT ACACTAGAAG AAATGATGAC AGCATGTCAG GGAGTGGGAG	1860
GACCTGGTCA TAAAGCAAGA GTTTGGCGG AAGCGATGAG CCAAGTAACA AATTCACTA	1920

CCATAATGAT GCAGAGAGGC AATTTAGGA ATCAAAGAAA GATTATCAAG TGCTTCATT 1980
GTGGCAAAGA AGGGCACATA GCCAAAATT GCAGGGCCCC TAGGAAAAGG GGCTGTTGGA 2040
AATGTGGAAA GGAAGGACAC CAAATGAAAG ATTGTACTGA GAGACAGGCT AATTTTTAG 2100
GGAAGATCTG GCCTTCCTGC AAGGGAAGGC AGGGAATTCT CCTCAGAGCA GAACAGAGCC 2160
AACAGCCCCA CCAGAAGAGA GCTTCAGGTT TGGGGAAGAG ACAACAACTC CCTATCAGAA 2220
GCAGGAGAAG AAGCAGGAGA CGATAGACAA GGACCTGTAT CCTTTAGCTT CCCTCAAATC 2280
ACTCTTGCG AACGACCCAT TGTCACAATA AAGATAGGGG GGCAACTAAA GGAAGCTCTA 2340
TTAGATACAG GAGCAGATGA TACAGTATTA GAAGAAATGA ATTTGCCAGG AAGATGGAAA 2400
CCAAAAATGA TAGGGGAAT TGGAGGTTTT ATCAAAGTAA GACAGTATGA TCAGATAACC 2460
ATAGAAATCT GTGGACATAA AGCTATAGGT ACAGTATTAG TAGGACCTAC ACCTGTCAAC 2520
ATAATTGGAA GAAATCTGTT GACTCAGCTT GGGTGCACCT TAAATTTCC CATTAGTCCT 2580
ATTGAAACTG TACCACTAAA ATAAAGCCA GGAATGGATG GCCCAAAAGT TAAACAATGG 2640
CCATTGACAG AAGAAAAAAT AAAAGCATTAA ATAGAAATTT GTACAGAAAT GGAAAAGGAA 2700
GGGAAAATTT CAAAAATTGG GCCTGAAAAT CCATACAATA CTCCAGTATT TGCCATAAAG 2760
AAAAAGACA GTACTAAATG GAGAAAATTA GTAGATTCA GAGAACTTAA TAAGAAAATC 2820
CAAGACTTCT GGGAAAGTTCA ATTAGGAATA CCACATCCTG CAGGGTTAAA AAAGAAAAAA 2880
TCAGTAACAG TACTGGATGT GGGTGATGCA TATTTTCAG TTCCCTTAGA TAAAGACTTC 2940
AGGAAGTATA CTGCATTAC CATACTAGT ATAAACAATG AAACACCAGG GATTAGATAT 3000
CAGTACAATG TGCTTCACA GGGATGGAAA GGATCACCAG CAATATTCCA AAGTAGCATG 3060
ACAAAAATCT TAGAGCCTTT TAGAAAACAA AATCCAGACA TAGTTATCTA TCAATACATG 3120
GATGATTGT ATGTAGGATC TGACTTAGAA ATAGGGCAGC ATAGAGCAAA AATAGAGGAA 3180
CTGAGACGAC ATCTGTTGAG GTGGGGATT ACCACACCAG ACAAAAAACAA TCAGAAAGAA 3240
CCTCCATTCC TTTGGATGGG TTATGAACTC CATCCTGATA AATGGACAGT ACAGCCTATA 3300
GTGCTGCCAG AAAAAAGACAG CTGGACTGTC AATGACATAC AGAAGTTAGT GGGAAAATG 3360
AATTGGCAA GTCAAATTAA CGCAGGGATT AAAGTAAAGC AATTATGTAA ACTCCTTAGA 3420
GGAACCAAAG CACTAACAGA AGTAATACCA CTAACAGAAG AAGCAGAGCT AGAACTGGCA 3480
GAAAACAGGG AAATTCTAAA AGAACCCAGTA CATGGAGTGT ATTATGACCC ATCAAAAGAC 3540
TTAATAGCAG AAGTACAGAA GCAGGGCAA GGCCAAATGGA CATATCAAAT TTATCAAGAG 3600
CCATTAAAAA ATCTGAAAAC AGGCAAATAT GCAAGAATGA GGGGTGCCCA CACTAATGAT 3660
GTAAAACAAT TAACAGAGGC AGTGCAAAAA ATAGCCACAG AAAGCATACT AATATGGGAA 3720
AAGACTCCTA AATTTAGACT ACCCATAACAA AAAGAAAACAT GGGAAACATG GTGGACAGAG 3780
TATTGGCAAG CCACCTGGAT TCCTGAGTGG GAGTTTGTCA ATACCCCTCC CTTAGTGAAA 3840
TTATGGTACC AGTTAGAGAA AGAACCCATA GTAGGAGCAG AAACCTTCTA TGTAGATGGG 3900
GCAGCTAACAA GGGAGACTAA AAAGGAAAAA GCAGGATATG TTACTAACAG AGGAAGACAA 3960

AAGGTTGTCT CCCTAACTGA CACAACAAAT CAGAAGACTG AGTTACAAGC AATTCACTA 4020
GCTTGCAAG ATTCAGGGTT AGAAGTAAAC ATAGTAACAG ACTCACAAATA TGCAATTAGGA 4080
ATCATTCAAG CACAACCAAGA TAAAAGTGA TCAGAGTTAG TCAGTCAAAT AATAGAGCAG 4140
TTAATAAAAA AGGAAAAGGT CTATCTGGCA TGGGTACCAAG CACACAAAGG AATTGGAGGA 4200
AATGAACAAG TAGATAAAATT AGTCAGTGCT GGAATCAGGA AAGTACTATT TTTAGATGGA 4260
ATAGATAAGG CCCAAGAAGA CCATGAGAAA TATCACAGTA ATTGGAGAGC AATGGCTAGT 4320
GACTTTAACCC TACCACCTAT AGTAGAAAA GAAATAGTAG CCAGCTGTGA TAAATGTCAG 4380
CTAAAAGGAG AAGCCATGCA TGGACAAGTA GACTGTAGTC CAGGAATATG GCAACTAGAT 4440
TGTACACATT TAGAAGGAAA AGTTATCCTG GTAGCAGTTC ATGTAGCCAG TGGATACATA 4500
GAAGCAGAAG TTATTCCAGC AGAGACAGGG CAGGAGACAG CATACTTTCT CTAAAATTA 4560
GCAGGAAGAT GGCCAGTAAA ACAAATACAT ACAGACAATG GCCCCAATTT CACCAAGTACT 4620
ACGGTTAAGG CCGCCTGTTG GTGGCGGGG ATCAAGCAGG AATTGGCAT TCCCTACAAT 4680
CCCCAAAGTC AAGGAGTAAT AGAATCTATG AATAAAGAAT TAAAGAAAAT TATAGGACAG 4740
GTAAGAGATC AGGCTGAACA TCTTAAGACA GCAGTACAAA TGGCACTATT CATCCACAAT 4800
TTTAAAAGAA AAGGGGGGAT TGGGGGGTAC AGTGCAGGGG AAGAAATAGT AGACATAATA 4860
GCAACAGACA TACAAACTAA AGAACTACAA AAACAAATTA CAAAAATTCA AAATTTCGG 4920
GTTTATTACA GGGACAGCAG AGATCCACTT TGGAAAGGAC CAGCAAAGCT TCTCTGGAAA 4980
GGTGAAGGGG CAGTAGTAAT ACAAGATAAT AGTGCACATAA AAGTAGTGCC AAGAAGAAAA 5040
GCAGGAGATCA TTAGGGATTA TGGAAAACAG ATGGCAGGTG ATGATTGTGT GGCAAGTAGA 5100
CAGGATGAGG ATTAGAACAT GGAAAAGTT AGTAAACAC CATATGTATA TTTCAAAGAA 5160
AGCTAAAGGA TGGTTTATA GACATCACTA TGAAAGCACT CATCCAGAA TAAGTTCAGA 5220
AGTACACATC CCACTAGGGG ATGCTAGATT GGTAAATAACA ACATATTGGG GTCTGCATAC 5280
AGGAGAAAGA GACTGGCATT TAGTCAGGG AGTCTCCATA GAATGGAGGA AAAAGAGATA 5340
TAGCACACAA GTAGACCCCTG ACCTAGCAGA CCACCTAATT CATCTGCATT ACTTTGATTG 5400
TTTTTCAGAC TCTGCCATAA GAAAGGCCAT ATTAGGACAT AGAGTTAGTC CTATTTGTGA 5460
ATTCAGCA GGACATAACA AGGTAGGATC TCTACAGTAC TTGGCACTAA CAGCATTAAAT 5520
AACACCAAAA AAGATAAAAGC CACCTTGCC TAGTGTAAAG AACTGACAG AGGATAGATG 5580
GAACAAGCCC CAGAAGACCA AGGGCCACAG AGGGAGCCAT ACAATCAATG GGCATTAGAG 5640
CTTTAGAGG AGCTTAAGAA TGAAGCTGTT AGACATTTC CTAGGATATG GCTCCATGGC 5700
TTAGGGCAAC ATATCTATGA AACTTATGGG GATACTTGGG CAGGAGTGGA AGCCATAATA 5760
AGAATTCTAC AACAACTGCT GTTTATTCAAT TTCAGAATTG GGTGTCGACA TAGCAGAATA 5820
GGCATTATTC GACAGAGGAG AGCAAGAAAT GGAGCCAGTA GATCCTAGAC TAGAGCCCTG 5880
GAAGCATCCA GGAAGTCAGC CTAAGACTGC TTGTACCACT TGCTATTGTA AAAAGTGTG 5940

CTTTCATTGC CAAGTTGTT TCACAAAAAA AGCCCTAGGC ATCTCCTATG GCAGGAAGAA 6000
 GCGGAGACAG CGACGAAGAG CTCCCTGAAGA CAGTCAGACT CATCAAGTTT CTCTACCAAA 6060
 GCAGTAAGTA GTACATGTAA TGCAACCTTT AGTAATAGCA GCAATAGTAG CATTAGTAGT 6120
 AGCAGGAATA ATAGCAATAG TTGTGTGATC CATACTTAC ATAGAATATA GGAAAATAAG 6180
 AAGACAAAGA AAAATAGACA GGGTAATTGA CAGAATAAGC GAAAGAGCAG AAGACAGTGG 6240
 CA ATG AGA GTG AAG GGG ATC AGG AGG AAT TAT CAG CAC TGG TGG GGA 6287
 Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly
 1 5 10 15
 TGG GGC ACG ATG CTC CTT GGG TTA TTA ATG ATC TGT AGT GCT ACA GAA 6335
 Trp Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu
 20 25 30
 AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT GTG TGG AAA GAA GCA 6383
 Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
 35 40 45
 ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA GCA TAT GAT ACA GAG 6431
 Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
 50 55 60
 GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA CCC ACA GAC CCC AAC 6479
 Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 65 70 75
 CCA CAA GAA GTA GAA TTG GTA AAT GTG ACA GAA AAT TTT AAC ATG TGG 6527
 Pro Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
 80 85 90 95
 AAA AAT AAC ATG GTA GAA CAG ATG CAT GAG GAT ATA ATC AGT TTA TGG 6575
 Lys Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
 100 105 110
 GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC CCA CTC TGT GTT ACT 6623
 Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 115 120 125
 TTA AAT TGC ACT GAT TTG AGG AAT ACT ACT AAT ACC AAT AAT AGT ACT 6671
 Leu Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Asn Ser Thr
 130 135 140
 GCT AAT AAC AAT AGT AAT AGC GAG GGA ACA ATA AAG GGA GGA GAA ATG 6719
 Ala Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met
 145 150 155
 AAA AAC TGC TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT AAG ATG CAG 6767
 Lys Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Lys Met Gln
 160 165 170 175
 AAA GAA TAT GCA CTT CTT TAT AAA CTT GAT ATA GTA TCA ATA AAT AAT 6815
 Lys Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asn Asn
 180 185 190
 GAT AGT ACC AGC TAT AGG TTG ATA AGT TGT AAT ACC TCA GTC ATT ACA 6863
 Asp Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr
 195 200 205
 CAA GCT TGT CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAC TAT TGT 6911
 Gln Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys
 210 215 220

GCC CCG GCT GGT TTT GCG ATT CTA AAG TGT AAC GAT AAA AAG TTC AGT Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser 225 230 235	6959
GGA AAA GGA TCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA Gly Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly 240 245 250 255	7007
ATT AGG CCA GTA GTA TCA ACT CAA CTG CTG TTA AAT GGC AGT CTA GCA Ile Arg Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala 260 265 270	7055
GAA GAA GAG GTA GTA ATT AGA TCT GAG AAT TTC AAT GAT AAT GCT AAA Glu Glu Glu Val Val Ile Arg Ser Glu Asn Phe Asn Asp Asn Ala Lys 275 280 285	7103
ACC ATC ATA GTA CAT CTG AAT GAA TCT GTA CAA ATT AAT TGT ACA AGA Thr Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg 290 295 300	7151
CCC AAC TAC AAT AAA AGA AAA AGG ATA CAT ATA GGA CCA GGG AGA GCA Pro Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala 305 310 315	7199
TTT TAT ACA ACA AAA AAT ATA ATA GGA ACT ATA AGA CAA GCA CAT TGT Phe Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys 320 325 330 335	7247
AAC ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AGA CAG ATA GTT AGC Asn Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser 340 345 350	7295
AAA TTA AAA GAA CAA TTT AAG AAT AAA ACA ATA GTC TTT AAT CAA TCC Lys Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser 355 360 365	7343
TCA GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA GGG Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly 370 375 380	7391
GAA TTT TTC TAC TGT AAT ACA TCA CCA CTG TTT AAT AGT ACT TGG AAT Glu Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn 385 390 395	7439
GGT AAT AAT ACT TGG AAT AAT ACT ACA GGG TCA AAT AAC AAT ATC ACA Gly Asn Asn Thr Trp Asn Asn Thr Thr Gly Ser Asn Asn Asn Ile Thr 400 405 410 415	7487
CTT CAA TGC AAA ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA Leu Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly 420 425 430	7535
AAA GCA ATA TAT GCC CCT CCC ATT GAA GGA CAA ATT AGA TGT TCA TCA Lys Ala Ile Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser 435 440 445	7583
AAT ATT ACA GGG CTA CTA TTA ACA AGA GAT GGT GGT AAG GAC ACG GAC Asn Ile Thr Gly Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp 450 455 460	7631
ACG AAC GAC ACC GAG ATC TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC Thr Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp 465 470 475	7679

AAT TGG AGA AGT GAA TTA TAT AAA TAT AAA GTA GTA ACA ATT GAA CCA	7727
Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro	
480 485 490 495	
TTA GGA GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAG AGA GAA	7775
Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu	
500 505 510	
AAA AGA GCA GCG ATA GGA GCT CTG TTC CTT GGG TTC TTA GGA GCA GCA	7823
Lys Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala	
515 520 525	
GGA AGC ACT ATG GGC GCA GCG TCA GTG ACG CTG ACG GTA CAG GCC AGA	7871
Gly Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg	
530 535 540	
CTA TTA TTG TCT GGT ATA GTG CAA CAG CAG AAC AAT TTG CTG AGG GCC	7919
Leu Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala	
545 550 555	
ATT GAG GCG CAA CAG CAT ATG TTG CAA CTC ACA GTC TGG GGC ATC AAG	7967
Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys	
560 565 570 575	
CAG CTC CAG GCA AGA ATC CTG GCT GTG GAA AGA TAC CTA AAG GAT CAA	8015
Gln Leu Gln Ala Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln	
580 585 590	
CAG CTC CTG GGG ATT TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC ACT	8063
Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr	
595 600 605	
ACT GTG CCT TGG AAT GCT AGT TGG AGT AAT AAA TCT CTG GAT GAT ATT	8111
Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile	
610 615 620	
TGG AAT AAC ATG ACC TGG ATG CAG TGG GAA AGA GAA ATT GAC AAT TAC	8159
Trp Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr	
625 630 635	
ACA AGC TTA ATA TAC TCA TTA CTA GAA AAA TCG CAA ACC CAA CAA GAA	8207
Thr Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu	
640 645 650 655	
ATG AAT GAA CAA GAA TTA TTG GAA TTG GAT AAA TGG GCA AGT TTG TGG	8255
Met Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp	
660 665 670	
AAT TGG TTT GAC ATA ACA AAT TGG CTG TGG TAT ATA AAA ATA TTC ATA	8303
Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile	
675 680 685	
ATG ATA GTA GGA GGC TTG GTA GGT TTA AGA ATA GTT TTT GCT GTA CTT	8351
Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu	
690 695 700	
TCT ATA GTG AAT AGA GTT AGG CAG GGA TAC TCA CCA TTG TCG TTG CAG	8399
Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln	
705 710 715	
ACC CGC CCC CCA GTT CCG AGG GGA CCC GAC AGG CCC GAA GGA ATC GAA	8447
Thr Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu	
720 725 730 735	

GAA GAA GGT GGA GAG AGA GAC AGA GAC ACA TCC GGT CGA TTA GTG CAT	8495																																																																																																									
Glu Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His																																																																																																										
740	745	750		GGA TTC TTA GCA ATT ATC TGG GTC GAC CTG CGG AGC CTG TTC CTC TTC	8543	Gly Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe		755	760	765		AGC TAC CAC CAC TTG AGA GAC TTA CTC TTG ATT GCA GCG AGG ATT GTG	8591	Ser Tyr His His Leu Arg Asp Leu Leu Ile Ala Ala Arg Ile Val		770	775	780		GAA CTT CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT	8639	Glu Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn		785	790	795		CTC CTA CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG	8687	Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu		800	805	810	815	CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735	Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile		820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746		
750																																																																																																										
GGA TTC TTA GCA ATT ATC TGG GTC GAC CTG CGG AGC CTG TTC CTC TTC	8543																																																																																																									
Gly Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe																																																																																																										
755	760	765		AGC TAC CAC CAC TTG AGA GAC TTA CTC TTG ATT GCA GCG AGG ATT GTG	8591	Ser Tyr His His Leu Arg Asp Leu Leu Ile Ala Ala Arg Ile Val		770	775	780		GAA CTT CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT	8639	Glu Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn		785	790	795		CTC CTA CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG	8687	Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu		800	805	810	815	CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735	Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile		820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746										
765																																																																																																										
AGC TAC CAC CAC TTG AGA GAC TTA CTC TTG ATT GCA GCG AGG ATT GTG	8591																																																																																																									
Ser Tyr His His Leu Arg Asp Leu Leu Ile Ala Ala Arg Ile Val																																																																																																										
770	775	780		GAA CTT CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT	8639	Glu Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn		785	790	795		CTC CTA CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG	8687	Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu		800	805	810	815	CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735	Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile		820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746																		
780																																																																																																										
GAA CTT CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT	8639																																																																																																									
Glu Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn																																																																																																										
785	790	795		CTC CTA CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG	8687	Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu		800	805	810	815	CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735	Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile		820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746																										
795																																																																																																										
CTC CTA CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGC TTG	8687																																																																																																									
Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu																																																																																																										
800	805	810	815	CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735	Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile		820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746																																		
810	815																																																																																																									
CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGG ACA GAT AGG GTT ATA	8735																																																																																																									
Leu Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile																																																																																																										
820	825	830		GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783	Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg		835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746																																										
830																																																																																																										
GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA	8783																																																																																																									
Glu Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg																																																																																																										
835	840	845		ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833	Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu		850	855		CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893			AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953			CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013			AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073			CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133			AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193			GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253			AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313			TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373			AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433			AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493			GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553			TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613			CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673			GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733			AAAATCTCTA GCA 9746																																																		
845																																																																																																										
ATA AGA CAG GGC TTG GAA AGG GCT TTG CTA TAAGATGGGT GGCAAATGGT	8833																																																																																																									
Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu																																																																																																										
850	855																																																																																																									
CRAAACGTGT GACTGGATGG CCTACTGTAA GGGAAAAAAT GAGACGAGCT GAACCAGCTG 8893																																																																																																										
AGCCAGCAGC AGATGGGGTG GGAGCAGCAT CCCGAGACCT GGAAAAACAT GGAGCACTCA 8953																																																																																																										
CAAGTAGCAA TACAGCAGCT ACCAATGCTG ATTGTGCCCTG GCTAGAAGCA CAAAGGGAGG 9013																																																																																																										
AGGAAGTGGG TTTTCCAGTC AGACCTCAGG TACCTTTAAG ACCAATGACT TACAAAGCAG 9073																																																																																																										
CTTTAGATCT TAGCCACTTT TTAAAAGAAA AGGGGGACT GGATGGGTTA ATTTACTCCC 9133																																																																																																										
AAAAGAGACA AGACATCCCTT GATCTGTGGG TCTACCACAC ACAAGGCTAC TTCCCTGATT 9193																																																																																																										
GGCAGAACTA CACACCAGGG CCAGGGATCA GATATCCACT GACCTTGGGA TGGTGCTTCA 9253																																																																																																										
AGCTAGTACC AGTTGAGCCA GAGAAGATAG AAGAGGCCAA TAAAGGAGAG AACAACTGCT 9313																																																																																																										
TGTTACACCC TATGAGCCAG CATGGGATGG ATGACCCGGGA GAGAGAAGTG TTAGTGTGGA 9373																																																																																																										
AGTCTGACAG CCACCTAGCA TTTCAGCATT ATGCCCGAGA GCTGCATCCG GAGTACTACA 9433																																																																																																										
AGAACTGCTG ACATCGAGCT ATCTACAAGG GACTTTCCGC TGGGGACTTT CCAGGGAGGT 9493																																																																																																										
GTGGCCTGGG CGGGACCGGG GAGTGGCGAG CCCTCAGATG CTGCATATAA GCAGCTGCTT 9553																																																																																																										
TCTGCCTGTA CTGGGTCTCT CTGGTTAGAC CAGATCTGAG CCTGGGAGCT CTCTGGCTAA 9613																																																																																																										
CTAGGGAACC CACTGCTTAA GCCTCAATAA AGCTTGCCTT GAGTGCTTCA AGTAGTGTGT 9673																																																																																																										
GCCCGTCTGT TATGTGACTC TGGTAGCTAG AGATCCCTCA GATCCTTTA GGCAGTGTGG 9733																																																																																																										
AAAATCTCTA GCA 9746																																																																																																										

Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp			
1	5	10	15
10	15		

Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu Lys
 20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
 35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
 50 55 60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80

Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp Lys
 85 90 95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125

Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Ser Thr Ala
 130 135 140

Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met Lys
 145 150 155 160

Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Lys Met Gln Lys
 165 170 175

Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asn Asn Asp
 180 185 190

Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln
 195 200 205

Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
 210 215 220

Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser Gly
 225 230 235 240

Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
 245 250 255

Arg Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala Glu
 260 265 270

Glu Glu Val Val Ile Arg Ser Glu Asn Phe Asn Asp Asn Ala Lys Thr
 275 280 285

Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro
 290 295 300

Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala Phe
 305 310 315 320

Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys Asn
 325 330 335

Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser Lys
 340 345 350

Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser Ser
 355 360 365

Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
 370 375 380
 Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly
 385 390 395 400
 Asn Asn Thr Trp Asn Asn Thr Thr Gly Ser Asn Asn Asn Ile Thr Leu
 405 410 415
 Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys
 420 425 430
 Ala Ile Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser Asn
 435 440 445
 Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp Thr
 450 455 460
 Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp Asn
 465 470 475 480
 Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu
 485 490 495
 Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
 500 505 510
 Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly
 515 520 525
 Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu
 530 535 540
 Leu Leu Ser Gly Ile Val Gln Gln Asn Asn Leu Leu Arg Ala Ile
 545 550 555 560
 Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln
 565 570 575
 Leu Gln Ala Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln
 580 585 590
 Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Thr
 595 600 605
 Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile Trp
 610 615 620
 Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr
 625 630 635 640
 Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu Met
 645 650 655
 Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn
 660 665 670
 Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met
 675 680 685
 Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser
 690 695 700
 Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr
 705 710 715 720

Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu
725 730 735

Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His Gly
740 745 750

Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser
755 760 765

Tyr His His Leu Arg Asp Leu Leu Ile Ala Ala Arg Ile Val Glu
770 775 780

Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu
785 790 795 800

Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu
805 810 815

Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu
820 825 830

Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg Ile
835 840 845

Arg Gln Gly Leu Glu Arg Ala Leu Leu
850 855

TABLE III

GATCAAGGGC CACAGAGGGA GCCACACAAT GAATGGACAC TAGAGCTTT AGAGGAGCTT	60
AAGAGTGAAG CTGTTAGACA CTTTCCTAGG ATATGGCTTC ATGGCTTAGG GCAACATATC	120
TATGAAACTT ATGGGGATAC TTGGGCAGGA GTGGAAGCCA TAATAAGAAT TCTGCAACAA	180
CTGCTGTTA TCCATTCAG GATTGGGTGC CAACATAGCA GAATAGGTAT TATTCAACAG	240
ACGAGAGCAA GAAATGGAGC CAGTAGATCC TAAACTAGAG CCCTGGAAGC ATCCAGGAAG	300
TCAGCCTAAG ACTGCTTGTA CCACTTGCTA TTGTAAAAAG TTGTTGCTTC ATTGCCAAGT	360
TTGCTTCATA ACAAAAGGCT TAGGCATCTC CTATGGCAGG AAGAAGCCGA GACAGCGACG	420
AAGAGCTCCT CAAGACAGTG AGACTCATCA AGTTTCTCTA TCAAAGCAGT AAGTAGTACA	480
TGTAATGCAA GCTTTACAAA TATCAGCTAT AGTAGGATTAA GTAGTAGCCAG CAATAATAGC	540
AATAGTTGTG TGGACCATAG TATTCACTAGA ATATAGGAAA ATATTAAGGC AAAGAAAAAT	600
AGACAGGTTA ATTGATAGAA TAACAGAAAG AGCAGAAGAC AGTGGCA ATG AGA GTG	656
Met Arg Val	
1	
ACG GAG ATC AGG AAG AGT TAT CAG CAC TGG TGG AGA TGG GGC ATC ATG	704
Thr Glu Ile Arg Lys Ser Tyr Gln His Trp Trp Arg Trp Gly Ile Met	
5	10
20	25
Leu Leu Gly Ile Leu Met Ile Cys Asn Ala Glu Glu Lys Leu Trp Val	30
35	
CTC CTT GGG ATA TTA ATG ATC TGT AAT GCT GAA GAA AAA TTG TGG GTC	752
Leu Leu Gly Ile Leu Met Ile Cys Asn Ala Glu Glu Lys Leu Trp Val	
40	45
ACA GTC TAT TAT GGG GTA CCT GTG TGG AAA GAA GCA ACC ACC ACT CTA	800
Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr Thr Thr Leu	
55	60
TTT TGT GCA TCA GAT CGT AAA GCA TAT GAT ACA GAG GTA CAT AAT GTT	848
Phe Cys Ala Ser Asp Arg Lys Ala Tyr Asp Thr Glu Val His Asn Val	
70	75
80	85
TGG GCC ACA CAT GCC TGT GTA CCC ACA GAC CCC AAC CCA CAA GAA GTA	896
Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro Gln Glu Val	
95	100
GAA TTG AAA AAT GTG ACA GAA AAT TTT AAC ATG TGG AAA AAT AAC ATG	944
Glu Leu Lys Asn Val Thr Glu Asn Phe Asn Met Trp Lys Asn Asn Met	
105	110
115	120
GTA GAA CAA ATG CAT GAG GAT ATA ATC AGT TTA TGG GAT CAA AGC CTA	992
Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp Gln Ser Leu	
125	130
135	140
145	
AAG CCA TGT GTA AAA TTA ACC CCA CTC TGT GTT ACT TTA AAT TGC ACT	1040
Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu Asn Cys Thr	
150	155
GAT TTG AGG AAT GCT ACT AAT GGG AAT GAC ACT AAT ACC ACT AGT AGT	1088
Asp Leu Arg Asn Ala Thr Asn Gly Asn Asp Thr Asn Thr Ser Ser	
160	165

AGC AGG GGA ATG GTG GGG GGA GGA GAA ATG AAA AAT TGC TCT TTC AAT	1136
Ser Arg Gly Met Val Gly Gly Glu Met Lys Asn Cys Ser Phe Asn	
150 155 160	
ATC ACC ACA AAC ATA AGA GGT AAG GTG CAG AAA GAA TAT GCA CTT TTT	1184
Ile Thr Thr Asn Ile Arg Gly Lys Val Gln Lys Glu Tyr Ala Leu Phe	
165 170 175	
TAT AAA CTT GAT ATA GCA CCA ATA GAT AAT AAT AGT AAT AAT AGA TAT	1232
Tyr Lys Leu Asp Ile Ala Pro Ile Asp Asn Asn Ser Asn Asn Arg Tyr	
180 185 190 195	
AGG TTG ATA AGT TGT AAC ACC TCA GTC ATT ACA CAG GCC TGT CCA AAG	1280
Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys	
200 205 210	
GTA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCC CCG GCT GGT TTT	1328
Val Ser Phe Glu Pro Ile Pro His Tyr Cys Ala Pro Ala Gly Phe	
215 220 225	
GCG ATT CTA AAG TGT AAA GAT AAG AAG TTC AAT GGA AAA GGA CCA TGT	1376
Ala Ile Leu Lys Cys Lys Asp Lys Lys Phe Asn Gly Lys Gly Pro Cys	
230 235 240	
ACA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA GTA GTA	1424
Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val	
245 250 255	
TCA ACT CAA CTG CTG TTA AAT GGC AGT CTA GCA GAA GAG GTA GTA	1472
Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val	
260 265 270 275	
ATT AGA TCC GCC AAT TTC GCG GAC AAT GCT AAA GTC ATA ATA GTA CAG	1520
Ile Arg Ser Ala Asn Phe Ala Asp Asn Ala Lys Val Ile Ile Val Gln	
280 285 290	
CTG AAT GAA TCT GTA GAA ATT AAT TGT ACA AGA CCC AAC AAC AAT ACA	1568
Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr	
295 300 305	
AGA AAA AGT ATA CAT ATA GGA CCA GGC AGA GCA TTT TAT ACA ACA GGA	1616
Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr Thr Gly	
310 315 320	
GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT AAC CTT AGT AGA GCA	1664
Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu Ser Arg Ala	
325 330 335	
AAA TGG AAT GAC ACT TTA AAT AAG ATA GTT ATA AAA TTA AGA GAA CAA	1712
Lys Trp Asn Asp Thr Leu Asn Lys Ile Val Ile Lys Leu Arg Glu Gln	
340 345 350 355	
TTT GGG AAT AAA ACA ATA GTC TTT AAG CAC TCC TCA GGA GGG GAC CCA	1760
Phe Gly Asn Lys Thr Ile Val Phe Lys His Ser Ser Gly Gly Asp Pro	
360 365 370	
GAA ATT GTG ACG CAC AGT TTT AAT TGT GGA GGG GAA TTT TTC TAC TGT	1808
Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys	
375 380 385	
AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG AAT GTT ACT GAA GAG TCA	1856
Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Val Thr Glu Glu Ser	
390 395 400	

AAT AAC ACT GTA GAA AAT AAC ACA ATC ACA CTC CCA TGC AGA ATA AAA Asn Asn Thr Val Glu Asn Asn Thr Ile Thr Leu Pro Cys Arg Ile Lys 405 410 415	1904
CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA AGA GCA ATG TAT GCC CCT Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met Tyr Ala Pro 420 425 430 435	1952
CCC ATC AGA GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG CTG CTA Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu 440 445 450	2000
TTA ACA AGA GAT GGT GGT CCT GAG GAC AAC AAG ACC GAG GTC TTC AGA Leu Thr Arg Asp Gly Gly Pro Glu Asp Asn Lys Thr Glu Val Phe Arg 455 460 465	2048
CCT GGA GGA GGA GAT ATG AGG GAT AAT TGG AGA AGT GAA TTA TAT AAA Pro Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys 470 475 480	2096
TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC AAG GCA Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala 485 490 495	2144
AAG AGA AGA GTG GTG CAG AGA GAA AAA AGA GCA GTG GGA ATA GGA GCT Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Ile Gly Ala 500 505 510 515	2192
GTG TTC CTT GGG TTC TTG GGA GCA GCA GGA AGC ACT ATG GGC GCA GCG Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala 520 525 530	2240
GCA ATG ACG CTG ACG GTA CAG GCC AGA CTA TTA TTG TCT GGT ATA GTG Ala Met Thr Leu Thr Val Gln Ala Arg Leu Leu Leu Ser Gly Ile Val 535 540 545	2288
CAA CAG CAG AAC AAT CTG CTG AGG GCT ATT GAG GCG CAA CAG CAT CTG Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu 550 555 560	2336
TTG CAA CTC ACA GTC TGG GGC ATC AAG CAG CTC CAG GCA AGA GTC CTG Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu 565 570 575	2384
GCT GTG GAA AGA TAC CTA AGG GAT CAA CAG CTC CTG GGG ATT TGG GGT Ala Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly 580 585 590 595	2432
TGC TCT GGA AAA CTC ATC TGC ACC ACT GCT GTG CCT TGG AAT GCT AGT Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser 600 605 610	2480
TGG AGT AAT AAA TCT CTG AAT AAG ATT TGG GAT AAC ATG ACC TGG ATA Trp Ser Asn Lys Ser Leu Asn Lys Ile Trp Asp Asn Met Thr Trp Ile 615 620 625	2528
GAG TGG GAC AGA GAA ATT AAC AAT TAC ACA AGC ATA ATA TAC AGC TTA Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser Ile Ile Tyr Ser Leu 630 635 640	2576
ATT GAA GAA TCG CAG AAC CAA CAA GAA AAG AAT GAA CAA GAA TTA TTA Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu 645 650 655	2624

GAA TTA GAT AAA TGG GCA AGT TTG TCG AAT TGG TTT GAC ATA ACA AAA Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Lys 660 665 670 675	2672
TGG CTG TGG TAT ATA AAA ATA TTC ATA ATG ATA GTA GGA GGC TTG ATA Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Ile 680 685 690	2720
GGT TTA AGA ATA GTT TTT TCT GTA CTT TCT ATA GTG AAT AGA GTT AGG Gly Leu Arg Ile Val Phe Ser Val Leu Ser Ile Val Asn Arg Val Arg 695 700 705	2768
CAG GGA TAC TCA CCA TTA TCG TTT CAG ACC CAC CTC CCA TCC TCG AGG Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro Ser Ser Arg 710 715 720	2816
GGA CCC GAC AGG CCC GGA GGA ATC GAA GAA GAA GGT GGA GAG AGA GAC Gly Pro Asp Arg Pro Gly Gly Ile Glu Glu Gly Gly Glu Arg Asp 725 730 735	2864
AGA GAC AGA TCC GGT CCA TTA GTG AAC GGA TTC TTG GCG CTT ATC TGG Arg Asp Arg Ser Gly Pro Leu Val Asn Gly Phe Leu Ala Leu Ile Trp 740 745 750 755	2912
GTC GAT CTG CGG AGC CTG TTC CTC AGC TAC CAC CGC TTG AGA GAC Val Asp Leu Arg Ser Leu Phe Leu Phe Ser Tyr His Arg Leu Arg Asp 760 765 770	2960
TTA CTC TTG ATT GTG ATG AGG ATT GTG GAA CTT CTG GGA CTA GCA GGG Leu Leu Leu Ile Val Met Arg Ile Val Glu Leu Leu Gly Leu Ala Gly 775 780 785	3008
GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT CTC CTA CAG TAT TGG AGT Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser 790 795 800	3056
CAG GAA CTA AAG AAT AGT GCT GTT AGC TTG CTC AAT GCC ACA GCT GTA Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Val 805 810 815	3104
GCA GTA GCT GAA GGG ACA GAT AGG GTT ATA GAA GTA TTA CAG AGA GCT Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Leu Gln Arg Ala 820 825 830 835	3152
GTT AGA GCT ATT CTC CAC ATA CCT AGA AGA ATA AGA CAG GGC TTG GAA Val Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu 840 845 850	3200
AGG GCT TTG CTA TAAGATGGGT GGCAAGTGGT CAAAAAGTAG TATAGTCGTA Arg Ala Leu Leu 855	3252
TGGCCTGCTG TAAGGAAAAG AATGAGAAGA ACTGAGCCAG CAGCAGATGG AGTAGGAGCA GTATCTAGAG ACCTGGAAAA ACATGGAGCA ATCACAAAGTA GCAATACAGC AGCTAACAAAT GCTGATTGTG CCTGGCTAGA AGCACAAAGAG GATGAAGAAG TGGGTTTCC AGTCAGACCT CAGGTACCTT TAAGACCAAT GACTCGCAGT GCAGCTATAG ATCTTAGCCA CTTTTTTAAG AAAAAGGGGG GACTGGAAGG GCTAATTCA CCCCCAAAAAA GACAAGATAT CCTTGATTTG TGGGTCTACC ACACACAGG CTACTTCCCT GATTGGCAGA ACTACACACC AGGGCCAGGG ACCAAGATTC CACTGACCTT TGGATGGTGC TTCAAGCTAG TACCAGTTGA GCCAGAGAAG	3492 3432 3552 3612 3672

GTAGAAGAGG CCAATGAAGG AGAGAACAAAC TGCTTGTAC ACCCTATGAG CCTGCATGGG 3732
 ATGGATGACC CGGAGAAAGA AGTGTAGCA TGGAAGTTG ACAGCAGCCT AGCATTCCAT 3792
 CACGTGGCCC GAGAA 3807
 Met Arg Val Thr Glu Ile Arg Lys Ser Tyr Gln His Trp Trp Arg Trp
 1 5 10 15
 Gly Ile Met Leu Leu Gly Ile Leu Met Ile Cys Asn Ala Glu Glu Lys
 20 25 30
 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
 35 40 45
 Thr Thr Leu Phe Cys Ala Ser Asp Arg Lys Ala Tyr Asp Thr Glu Val
 50 55 60
 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80
 Gln Glu Val Glu Leu Lys Asn Val Thr Glu Asn Phe Asn Met Trp Lys
 85 90 95
 Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110
 Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125
 Asn Cys Thr Asp Leu Arg Asn Ala Thr Asn Gly Asn Asp Thr Asn Thr
 130 135 140
 Thr Ser Ser Ser Arg Gly Met Val Gly Gly Glu Met Lys Asn Cys
 145 150 155 160
 Ser Phe Asn Ile Thr Thr Asn Ile Arg Gly Lys Val Gln Lys Glu Tyr
 165 170 175
 Ala Leu Phe Tyr Lys Leu Asp Ile Ala Pro Ile Asp Asn Asn Ser Asn
 180 185 190
 Asn Arg Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala
 195 200 205
 Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro
 210 215 220
 Ala Gly Phe Ala Ile Leu Lys Cys Lys Asp Lys Lys Phe Asn Gly Lys
 225 230 235 240
 Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg
 245 250 255
 Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala Glu Glu
 260 265 270
 Glu Val Val Ile Arg Ser Ala Asn Phe Ala Asp Asn Ala Lys Val Ile
 275 280 285
 Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
 290 295 300
 Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr
 305 310 315 320

Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu
 325 330 335
 Ser Arg Ala Lys Trp Asn Asp Thr Leu Asn Lys Ile Val Ile Lys Leu
 340 345 350
 Arg Glu Gln Phe Gly Asn Lys Thr Ile Val Phe Lys His Ser Ser Gly
 355 360 365
 Gly Asp Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly Glu Phe
 370 375 380
 Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Val Thr
 385 390 395 400
 Glu Glu Ser Asn Asn Thr Val Glu Asn Asn Thr Ile Thr Leu Pro Cys
 405 410 415
 Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met
 420 425 430
 Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr
 435 440 445
 Gly Leu Leu Leu Thr Arg Asp Gly Gly Pro Glu Asp Asn Lys Thr Glu
 450 455 460
 Val Phe Arg Pro Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
 465 470 475 480
 Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro
 485 490 495
 Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly
 500 505 510
 Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
 515 520 525
 Gly Ala Ala Ala Met Thr Leu Thr Val Gln Ala Arg Leu Leu Leu Ser
 530 535 540
 Gly Ile Val Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln
 545 550 555 560
 Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
 565 570 575
 Arg Val Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu Gly
 580 585 590
 Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp
 595 600 605
 Asn Ala Ser Trp Ser Asn Lys Ser Leu Asn Lys Ile Trp Asp Asn Met
 610 615 620
 Thr Trp Ile Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser Ile Ile
 625 630 635 640
 Tyr Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln
 645 650 655
 Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp
 660 665 670

Ile Thr Lys Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly
675 680 685

Gly Leu Ile Gly Leu Arg Ile Val Phe Ser Val Leu Ser Ile Val Asn
690 695 700

Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro
705 710 715 720

Ser Ser Arg Gly Pro Asp Arg Pro Gly Gly Ile Glu Glu Glu Gly Gly
725 730 735

Glu Arg Asp Arg Asp Arg Ser Gly Pro Leu Val Asn Gly Phe Leu Ala
740 745 750

Leu Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser Tyr His Arg
755 760 765

Leu Arg Asp Leu Leu Ile Val Met Arg Ile Val Glu Leu Leu Gly
770 775 780

Leu Ala Gly Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu Gln
785 790 795 800

Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala
805 810 815

Thr Ala Val Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Leu
820 825 830

Gln Arg Ala Val Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln
835 840 845

Gly Leu Glu Arg Ala Leu Leu
850 855

WHAT IS CLAIMED IS:

1. A substantially pure preparation of a molecular clone capable of yielding after transfection into recipient cells active cultures of the Human Immunodeficiency Virus Type 1 (HIV-1) virus strain MN-ST1, having the identifying characteristics of ATCC 40889.
2. A substantially pure preparation of DNA containing the envelope and rev coding sequences of the (HIV-1) virus strain BA-L, having the identifying characteristics of ATCC 40890.
3. A DNA segment encoding an envelope (env) protein of MN-ST1.
4. The DNA segment according to claim 3 having the sequence given in Table III.
5. A DNA segment encoding an env protein of BA-L.
6. A DNA segment according to claim 5 having the sequence given in Table III.
7. A purified MN-ST1 env protein.
8. The protein according to claim 7 having the sequence given in Table II.
9. A purified BA-L protein.
10. The protein according to claim 9 having the sequence given in Table III.
11. A DNA construct comprising:
 - i) the DNA segment according to claim 3;
 - and
 - ii) a vector.
12. The DNA construct according to claim 11 further comprising a DNA segment encoding a rev protein and a rev-responsive region.
13. A DNA construct comprising:
 - i) the DNA segment according to claim 5;
 - and
 - ii) a vector.
14. The DNA construct according to claim 13 further comprising a DNA segment encoding a rev protein and a rev-responsive region.

15. A recombinantly produced MN-ST1 env protein.
16. A recombinantly produced BA-L env protein.
17. A host cell stably transformed with said recombinant DNA construct according to claim 11 or claim 13, in a manner allowing expression of said viral protein encoded in said recombinant DNA molecule.
18. A method of producing a recombinant HIV-1 virus strain MN-ST1 protein comprising culturing said host cells according to claim 17, in a manner allowing expression of said viral protein and isolating said viral protein.
19. A vaccine for mammals against HIV-1 infection comprising a non-infectious antigenic portion of said MN-ST1 virus strain according to claim 1, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
20. A vaccine for mammals against HIV-infection comprising a non-infectious antigenic portion of said BA-L virus strain according to claim 2 in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
21. The vaccine according to claim 19 or claim 20 which further comprises an adjuvant.
22. A vaccine for mammals against HIV-1 infection comprising at least 5 amino acids of a MN-ST1 virus strain env protein, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
23. A vaccine for mammals against HIV-1 infection comprising at least 5 amino acids of a BA-L virus strain env protein, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
24. The vaccine according to claim 22 or 23 wherein said protein is a recombinantly produced protein.
25. A method of testing candidate vaccines against HIV-1 infection comprising administering said vaccine and the MN-ST1 virus strain according to claim 1,

to a test mammal and detecting the presence or absence of said infection.

26. A method of screening drugs for their ability to effect HIV-1 activity comprising contacting host cells according to claim 17, with said drug under conditions such that said activity of said virus can be effected.

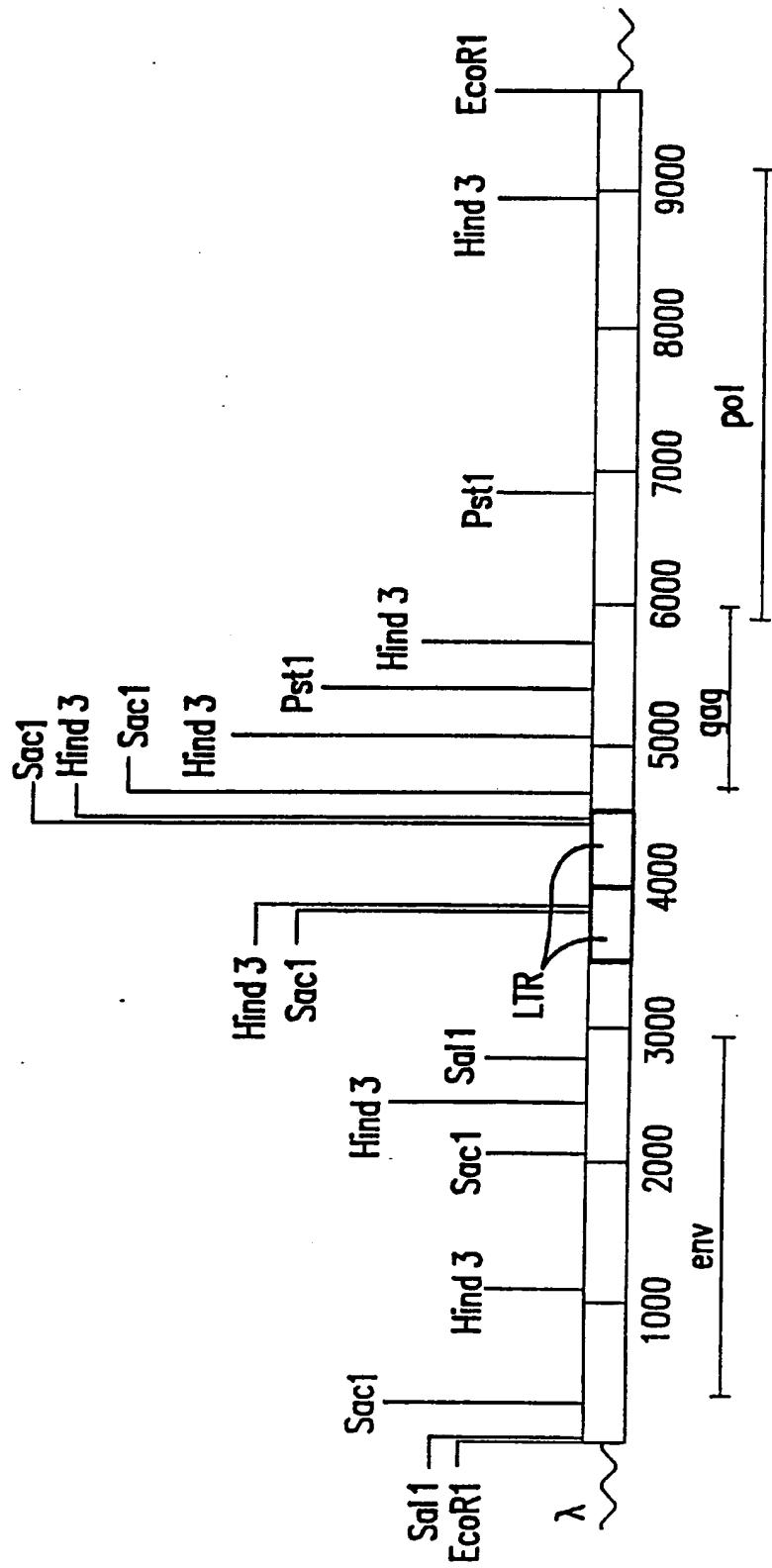
27. A bioassay for the detection of HIV-1 in a biological sample comprising the steps of:

i) coating a surface with at least 5 amino acids of a env protein from MN-ST1 or BA-L virus;

ii) contacting said coated surface with said sample; and

iii) detecting the presence or absence of a complex formed between said protein and antibodies specific therefor present in said sample.

1/7



SUBSTITUTE SHEET

FIG. 1

2/7

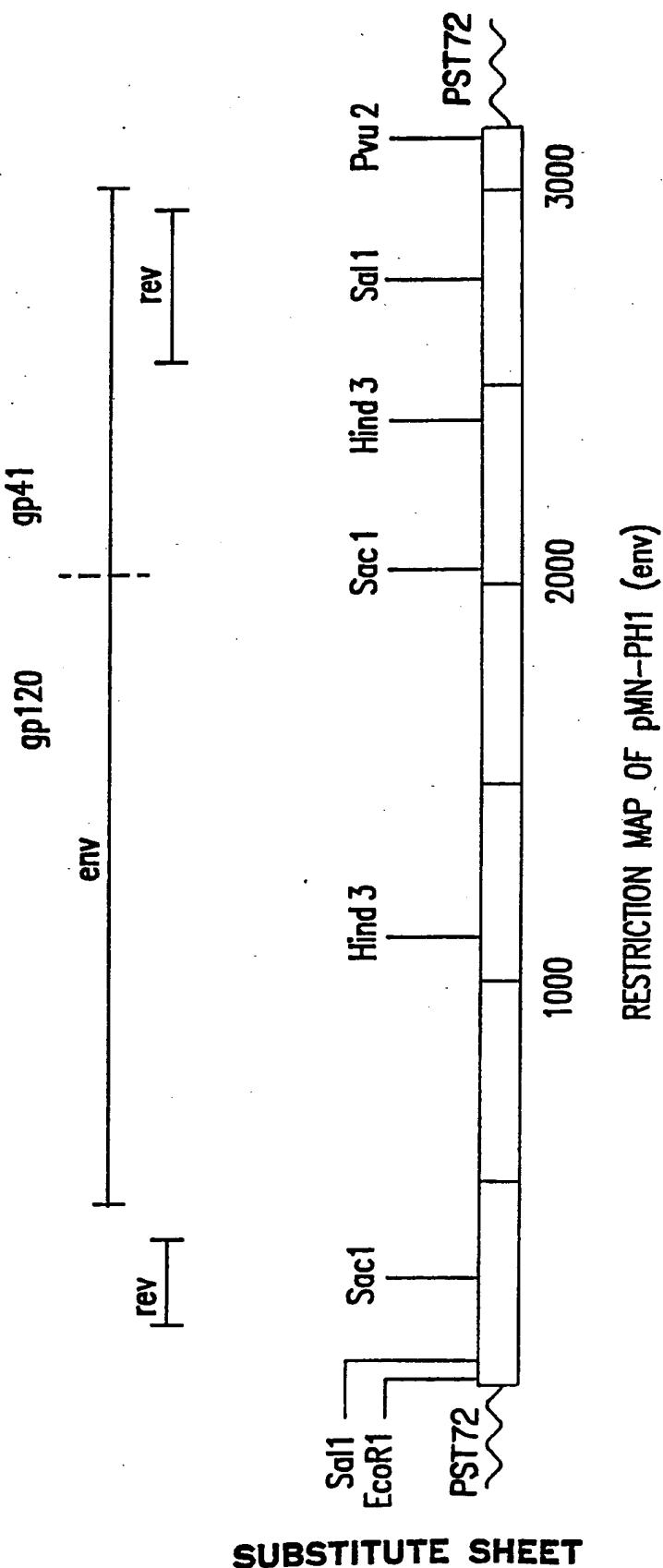
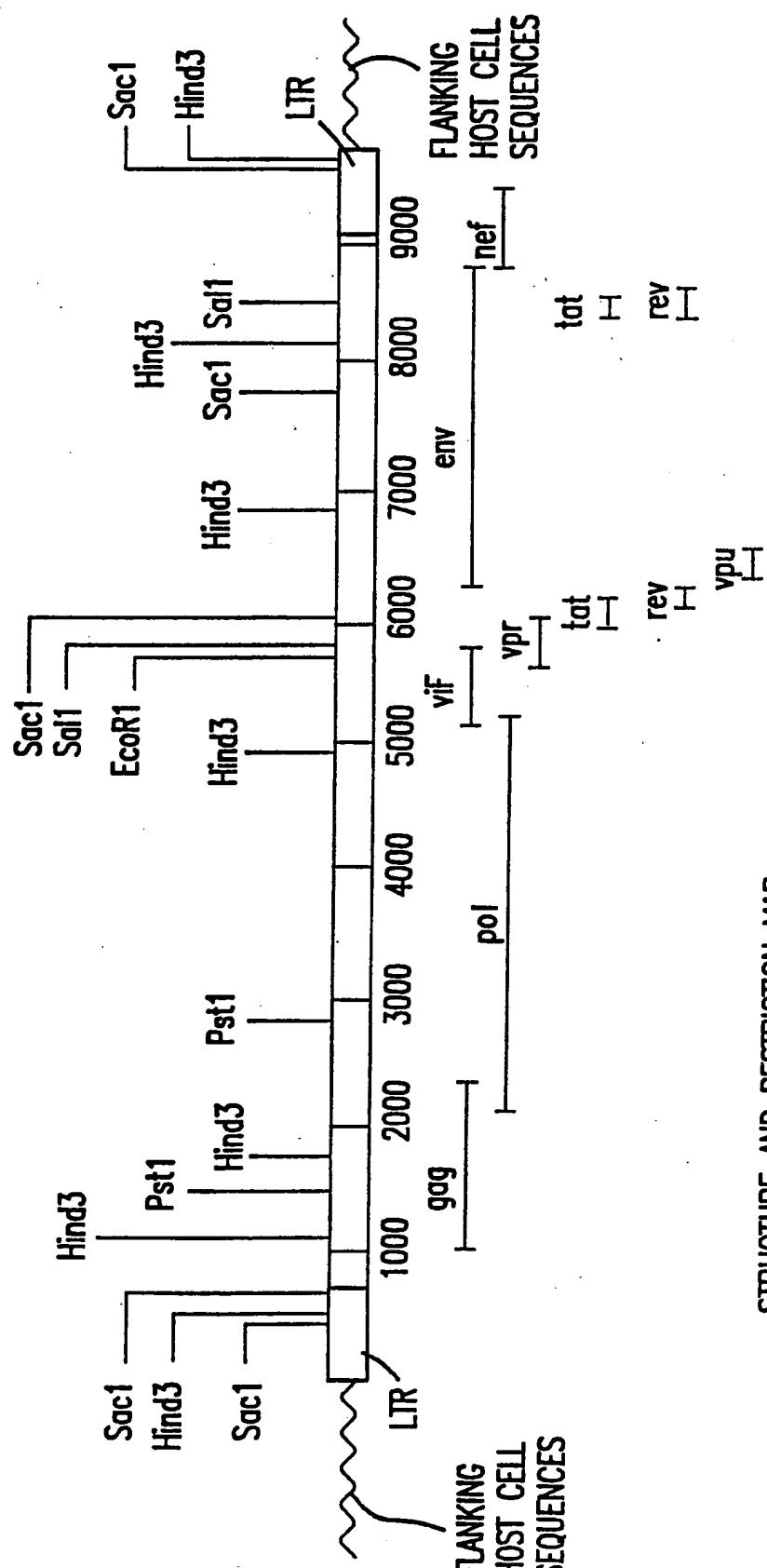


FIG. 2



SUBSTITUTE SHEET

FIG.3

4/7

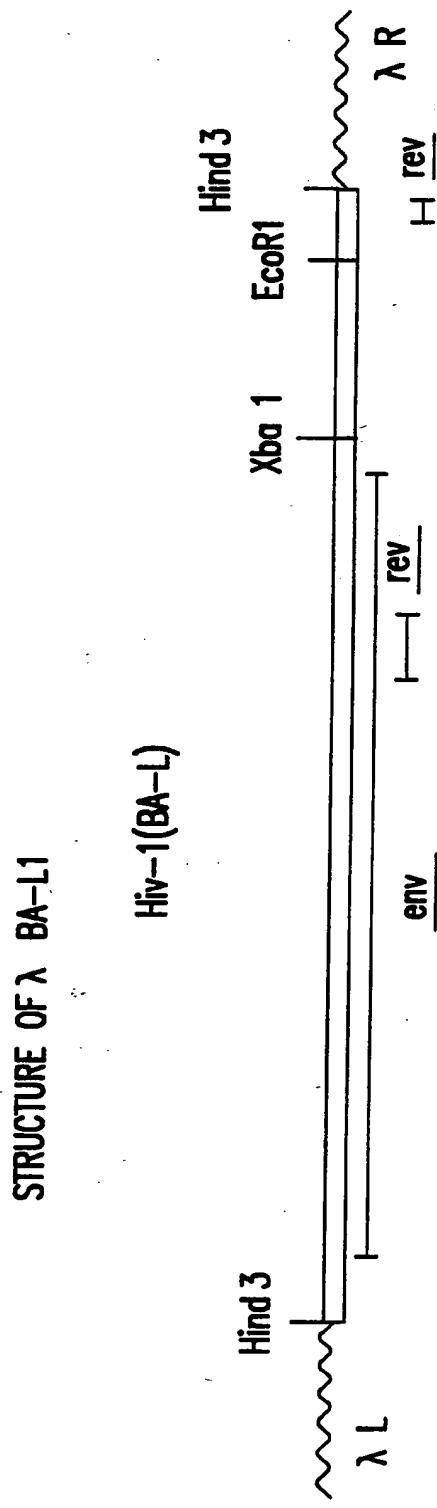


FIG.4

SUBSTITUTE SHEET

5/7

RESTRICTION MAP OF pBA-L1

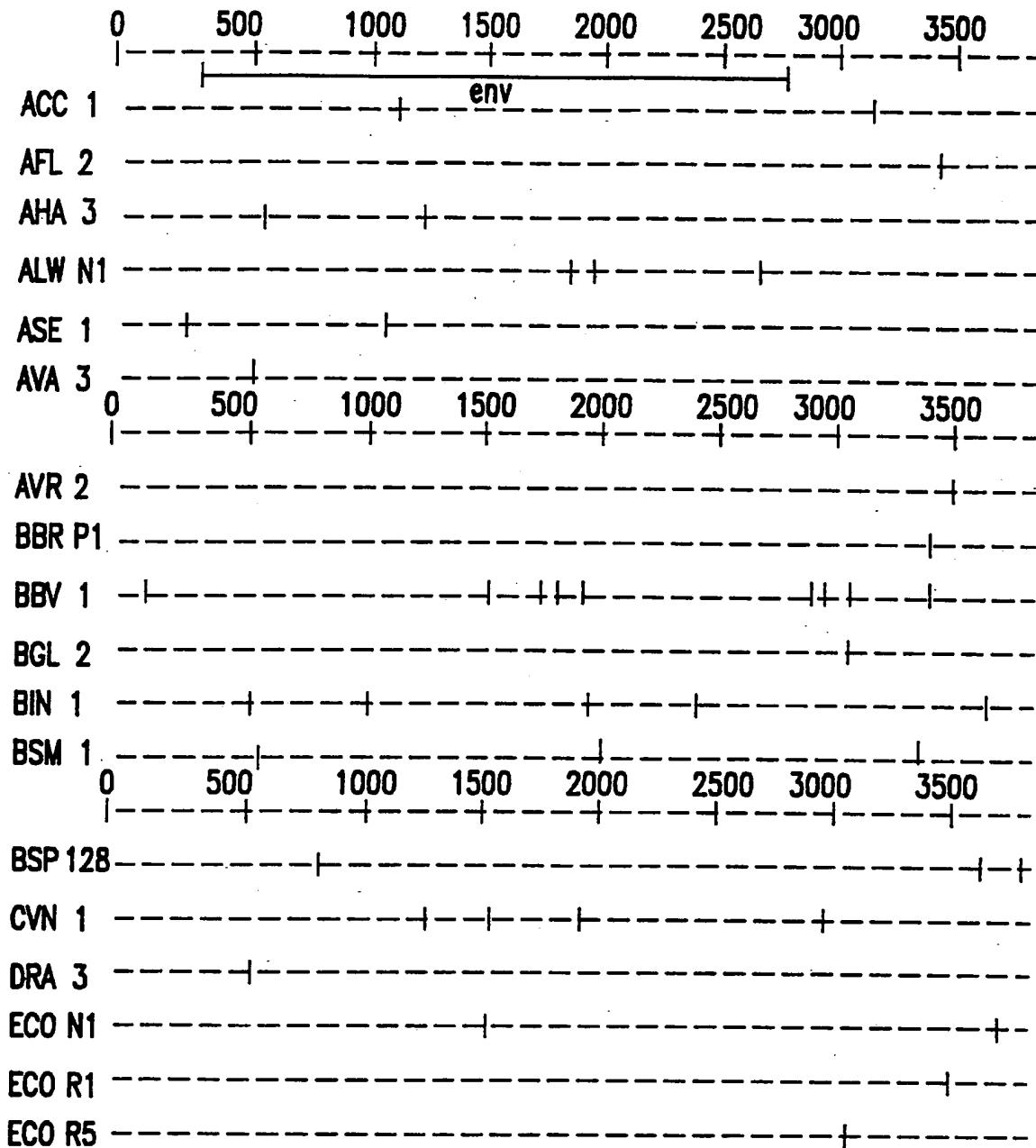
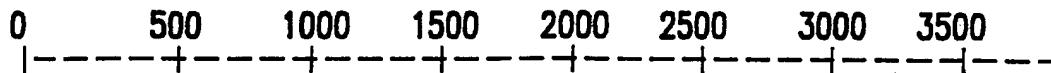


FIG.5A

SUBSTITUTE SHEET

6/7

BAFINAL



GCOT22 -----+-----

2

FOK 1 -----+-----+-----+-----+-----+-----

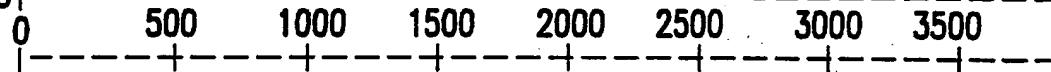
2

HAE 3 -----+-----+-----+-----+-----+-----+-----

HGA 1 -----+-----+-----

HHA 1 -----+-----+-----+-----

HIND 3 -----+-----+-----+-----+-----+-----+-----



HINF1 -----+-----+-----+-----+-----+-----+-----+-----

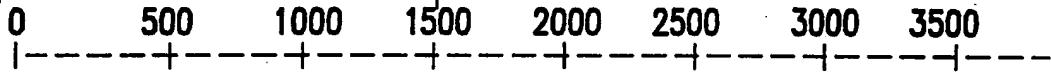
HPA 2 -----+-----+-----+-----+-----+-----+-----

HPH 2 -----+-----+-----

KPN 2 -----+-----+-----

Ksp632 -----+-----+-----+-----+-----+-----+-----

MAM 1 -----+-----+-----+-----+-----+-----+-----+-----



2

2

MBO 2 -----+-----+-----+-----+-----+-----+-----+-----+-----

MST 2 -----+-----+-----+-----+-----

NDE 1 -----+-----

NSI 1 -----+-----

PFL M1 -----+-----

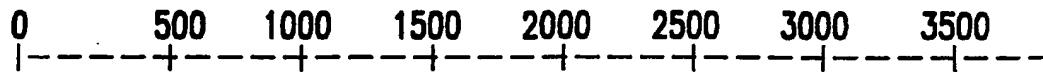
PML 1 -----+-----

FIG.5B

SUBSTITUTE SHEET

7/7

BAFINAL



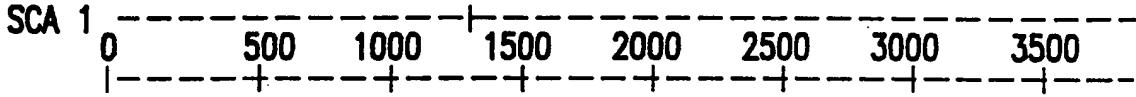
PVU 2 -----+

RRU 1 -----+

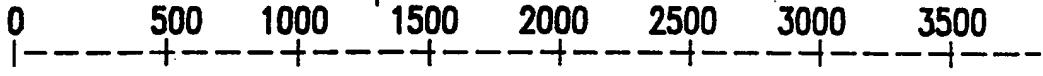
SAC 1 -----+

SAU 1 -----+

SAU 3A -+ + + +-----+



SCA 1 -----+



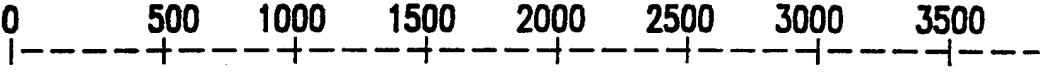
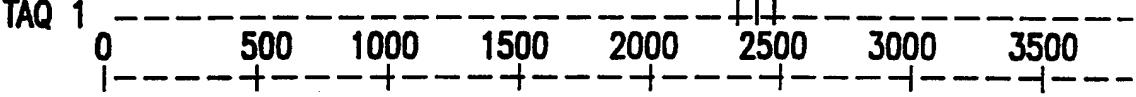
SFA N1 -----+-----+-----+

SNA 1 -----+

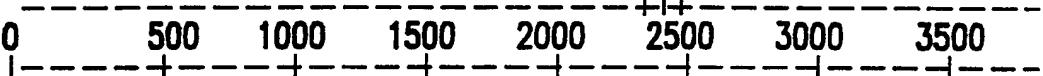
SPE 1 -----+

SSP 1 -+-----+-----+-----+

STU 1 -----+



TAQ 1 -----+-----+-----+-----+



TTH111 -----+-----+-----+-----+

XBA 1 -----+-----

XCM 1 -----+-----

XHO 1 -----+-----

XHO 2 -----+-----+-----

XMN 1 -----+-----

FIG.5C

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/07611

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC

**IPC(5): C07H 15/12; C12N 5/10, 7/02, 7/04, 15/49;
C07K 3/12, 13/00, 17/00; C12Q 1/70; A61K 39/10; G01N 33/53**

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System	Classification Symbols
U. S.	435/7.1, 235.1, 236, 240.1; 530/350; 536/27; 424/88

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸

**DIALOG DATABASES: BIOSIS PREVIEWS 1985+, MEDLINE 1975+,
NTIS, AIDSLINE, CA SEARCH, BIOTECHNOLOGY ABSTRACTS 1982+**

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No ¹³
Y	Science, Vol. 241, issued 22 July 1988, W. C. Koff et al, "Development and Testing of AIDS Vaccines," pages 426-432. See entire article.	1-24, 27
Y	Nature, Vol. 312, issued 20/27 December 1984, P. A. Luciw et al, "Molecular cloning of AIDS-associated retrovirus," pages 760-763. See entire article.	1-24, 27
Y	Science, Vol. 226, issued 07 December 1984, G. M. Shaw et al, "Molecular Characterization of Human T-Cell Leukemia (Lymphotropic) Virus Type III in the Acquired Immune Deficiency Syndrome," pages 1165-1171. See entire document.	1-24, 27
Y	Nature, Vol. 312, issued 20/27 December 1984, M. Alizon et al, "Molecular cloning of lymphadenopathy-associated virus," pages 757-760. See entire article.	1-24, 27

* Special categories of cited documents ¹⁰

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search:

15 JANUARY 1991

Date of Mailing of this International Search Report

30 JAN 1992

International Searching Authority

ISA/US

Signature of Authorized Officer

Johnny F. Railey II
JOHNNY F. RAILY II

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	Journal of Medical Virology, Vol. 19, issued 1986, H. Rübsamen-Waigmann et al, "Isolation of Variants of Lymphocytopathic Retroviruses From the Peripheral Blood and Cerebrospinal Fluid of Patients With ARC or AIDS," pages 335-344. See entire article.	1-24, 27
Y	Nature, Vol. 313, issued 24 January 1985, L. Ratner et al, "Complete nucleotide sequence of the AIDS virus, HTLV-III," pages 277-284. See entire article.	1-24, 27

V. OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons

1 Claim numbers _____ because they relate to subject matter¹² not required to be searched by this Authority, namely:

2 Claim numbers 25, 26 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out¹³, specifically:

Claims 25 and 26 are so vague and indefinite as to prevent a meaningful and thorough search.

3. Claim numbers _____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING²

This International Searching Authority found multiple inventions in this international application as follows:

See attachment

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

The additional search fees were accompanied by applicant's protest.

No protest accompanied the payment of additional search fees.

Attachment to Form PCT/ISA/210, Part VI
Continuation of OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group I: Claims 1 and 25, drawn to a first product, cloned HIV-1 strain MN-ST1, and the first appearing use of the product, a method of testing vaccines against HIV-1 using strain MN-ST1.

Group II: Claims 2, 5, 6, 13 and 14, drawn to a second product, HIV-1 strain BA-L env and rev coding sequences, DNA segments encoding the env gene, and vector constructs containing these sequences.

Group III: Claims 3, 4, 11 and 12, drawn to a third product, DNA encoding strain MN-ST1 env gene and vectors containing this env gene.

Group IV: Claim 17 (first species), drawn to a fourth product, host cells stably transformed with recombinant construct of claim 11.

Group V: Claim 17 (second species), drawn to a fifth product, host cells stably transformed with recombinant construct of claim 13.

Group VI: Claim 18 (first species), drawn to a method of use of the fourth product, host cells transformed with the recombinant construct of claim 11.

Group VII: Claim 18 (second species), drawn to a method of use of the fifth product, host cells transformed with the recombinant construct of claim 13.

Group VIII: Claims 7, 8 and 15, drawn to a sixth product, HIV-1 strain MN-ST1 env protein.

Group IX: Claims 9, 10 and 16, drawn to a seventh product, HIV-1 strain BA-L env protein.

Group X: Claims 19 and 21 (first species), drawn to an eighth product, vaccines using MN-ST1.

Group XI: Claims 20 and 21 (second species), drawn to a ninth product, vaccines using BA-L.

Attachment to Form PCT/ISA/210, Part VI
Continuation of OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group XII: Claims 22 and 24, drawn to a tenth product, vaccines using at least 5 amino acids of the env protein of MN-ST1.

Group XIII: Claim 23, drawn to an eleventh product, vaccines using at least 5 amino acids of the env protein of BA-L.

Group XIV: Claim 26, drawn to a twelfth product, a method of screening for drugs affecting HIV-1 activity.

Group XV: Claim 27, drawn to a thirteenth product, a bioassay to detect HIV-1 in biological samples.

The claims of Group I are drawn to a first product and a first specific method of use of the first product. Groups II-XV are drawn to separate products and methods of use of the products. PCT Rules 13.1 and 13.2 do not provide for multiple products and methods within a single general inventive concept. Note also 37 CFR § 1.475.

III DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Cell, Vol. 40, issued January 1985, S. Wain-Hobson et al, "Nucleotide Sequence of the AIDS Virus, LAV," pages 9-17. See entire article.	1-24, 27
Y	Science, Vol. 227, issued 01 February 1985, R. Sanchez-Pescador et al, "Nucleotide Sequence and Expression of an AIDS-Associated Retrovirus (ARV-2)," pages 484-492. See entire article.	1-24, 27
Y	Nature, Vol. 313, issued 07 February 1985, M. A. Huesing et al, "Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus," pages 450-458. See entire article.	1-24, 27
Y	Nature, Vol. 320, issued 10 April 1986, S.-L. Hu et al, "Expression of AIDS virus envelope gene in recombinant vaccinia viruses," pages 537-540. See entire article.	2-24, 27
Y	Nature, Vol. 320, issued 10 April 1986, S. Chakrabarti et al, "Expression of the HTLV-III envelope gene by a recombinant vaccinia virus," pages 535-537. See entire article.	2-24, 27
Y	Bio/Technology, Vol. 3, issued October 1985, T. W. Chang et al, "Detection of Antibodies to Human T-Cell Lymphotropic Virus-III (HTLV-III) with an Immunoassay Employing a Recombinant <u>Escherichia coli</u> -Derived Viral Antigenic Peptide," pages 905-909. See entire article.	2-24, 27
Y	Proc. Natl. Acad. Sci. USA, Vol. 84, issued October 1987, J. R. Rusche et al, "Humoral immune response to the entire human immunodeficiency virus envelope glycoprotein made in insect cells," pages 6924-6928. See entire article.	2-24, 27
Y	J. Virology, Vol. 63, No. 3, issued March 1989, M. Hadzopoulou-Cladaras et al, "The <u>rev</u> (<u>trs/art</u>) Protein of Human Immunodeficiency Virus Type 1 Affects Viral mRNA and Protein Expression via a <u>cis</u> -Acting Sequence in the <u>env</u> Region," pages 1265-1274. See entire article.	2, 7, 9, 12, 14, 15, 16, 20, 21

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	J. Virology, Vol. 64, No. 9, issued September 1990, P. J. Dillon et al, "Function of the Human Immunodeficiency Virus Types 1 and 2 Rev Proteins Is Dependent on Their Ability To Interact with a Structured Region Present in <u>env</u> Gene mRNA," pages 4428-4437. See entire article.	2, 7, 9, 12, 14, 15, 16, 20, 21
Y	Cell, Vol. 45, issued 06 June 1986, B. R. Starcich et al, "Identification and Characterization of Conserved and Variable Regions in the Envelope Gene of HTLV-III/LAV, the Retrovirus of AIDS," pages 637-648. See entire article.	2-24, 27
Y	J. Virology, Vol. 61, No. 2, issued February 1987, S. Modrow et al, "Computer-Assisted Analysis of Envelope Protein Sequences of Seven Human Immunodeficiency Virus Isolates: Prediction of Antigenic Epitopes in Conserved and Variable Regions," pages 570-578. See entire article.	2-24, 27
Y	Analytical Biochemistry, Vol. 151, issued 1985, D. Pauletti et al, "Application of a Modified Computer Algorithm in Determining Potential Antigenic Determinants Associated with the AIDS Virus Glycoprotein," pages 540-546. See entire article.	2-24, 27
Y	Virology, Vol. 164, issued 1988, C. Gурго et al, "Envelope Sequences of Two New United States HIV-1 Isolates," pages 531-536. See entire article.	2-24, 27
Y	J. Virology, Vol. 64, No. 5, issued May 1990, A. Aldovini et al, "Mutations of RNA and Protein Sequences Involved in Human Immunodeficiency Virus Type 1 Packaging Result in Production of Noninfectious Virus," pages 1920-1926. See entire article.	19

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.